

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: November 25, 2003, 04:55:18 ; Search time 3301 Seconds

(without alignments)  
10705.455 Million cell updates/sec

Title: US-10-021-368-2

Perfect score: 1454  
Sequence: 1 ACCAGCGGACAGACAGGCGC.....GTGCTGAAAAAAAAAAAA 1454

Scoring table:  
IDENTITY NUC  
Gapop 10.0, Gapex 1.0

Searched: 22781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estdb:\*  
2: em\_estdb:\*  
3: em\_estdb:\*  
4: em\_estdb:\*  
5: em\_estdb:\*  
6: em\_estdb:\*  
7: em\_estdb:\*  
8: em\_estdb:\*  
9: em\_estdb:\*  
10: em\_estdb:\*  
11: em\_estdb:\*  
12: em\_estdb:\*  
13: em\_estdb:\*  
14: em\_estdb:\*  
15: em\_estdb:\*  
16: em\_estdb:\*  
17: em\_estdb:\*  
18: em\_estdb:\*  
19: em\_estdb:\*  
20: em\_estdb:\*  
21: em\_estdb:\*  
22: em\_estdb:\*  
23: em\_estdb:\*  
24: em\_estdb:\*  
25: em\_estdb:\*  
26: em\_estdb:\*  
27: em\_estdb:\*  
28: em\_estdb:\*  
29: em\_estdb:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795.8	54.7	1140	12	BM809901 AGENCOURT
2	699.6	48.1	734	14	CA313179 UI-CF-FNO
3	687.4	47.3	714	12	BM019631 603647895
4	679.2	46.7	758	10	BM678912 602624621

5	672.6	46.3	794	10	BM469321
6	669.6	46.1	717	10	BM898352
7	652.8	44.9	686	12	BM983288
8	652.8	44.9	793	12	BM108513
9	650.4	44.7	686	10	BM679444
10	620.8	42.7	625	12	BM726274
11	619.2	42.6	736	9	BM859357
12	601	41.3	866	10	BM899455
13	600.4	41.3	614	14	BM216785
14	591.8	40.7	1116	10	BM736690
15	580	39.9	727	9	BM859383
16	566.6	39.0	592	9	BM601784
17	542.4	37.3	555	14	BM850946
18	511.4	35.2	525	12	BM783960
19	508.4	35.0	532	9	BM465016
20	507.8	34.9	577	12	BM837882
21	506.6	34.8	579	10	BM734253
22	506.6	34.8	507	13	BM837824
23	499	34.3	499	10	BM837896
24	492.4	33.9	505	12	BM759571
25	483.8	33.3	510	12	BM839092
26	483.4	33.2	739	12	BM108577
27	481.2	33.1	1338	11	AK009000
28	479.6	33.0	978	10	BM737427
29	474	32.6	897	13	BM517778
30	470.6	32.4	826	12	BM411732
31	451.4	31.0	951	13	BM920035
32	448.4	30.8	454	9	BM477256
33	447.4	30.8	449	9	BM459626
34	446	30.7	457	12	BM717931
35	445.4	30.6	470	9	BM290581
36	445.2	30.6	758	12	BM830577
37	438.8	30.2	549	12	BM830577
38	438.4	30.2	468	9	BM429058
39	437.4	30.1	452	9	BM478920
40	436.2	30.0	883	12	BM410819
41	432.2	29.7	437	12	BM106367
42	430.6	29.6	452	12	BM785939
43	423.6	29.1	520	9	BM102333
44	416.6	28.7	428	10	BM837644
45	414.4	28.5	440	9	BM554385

## ALIGNMENTS

RESULT 1  
LOCUS BM809901  
DEFINITION AGENCOURT 6581105 NIH\_MGC\_98 Homo sapiens cdna clone IMAGE:5454819  
ACCESSION BM809901  
VERSION 5', mRNA sequence.  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
REFERENCE  
1 (bases 1 to 1140)  
Mammalia, Eutheria, Chordata, Craniata, Vertebrata, Euteleostomi;  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strussberg, Ph.D.  
Email: cgsabbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNU)  
DNA sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNU at:  
http://image.llnl.gov  
Plate: LINC1950 row: 5 column: 04  
High quality sequence start: 17



6hr to LPS 24h  
TAG SEQ=CTGCTCAGT"  
BASE COUNT 182 a 159 c 222 g 171 t

Query Match 48.1%; Score 699.6; DB 14; Length 714;  
Best Local Similarity 99.2%; Pred. No. 7,66-131;  
Matches 724; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

727 GCTGAGTGGACGGGGGCGAGAGACCTTGGCCAGAGTGAAGTGGAGAGCCCTGGTCTGT 786  
734 GTTGAGTACACGGGGGCGAGAGACCTTGGCCAGAGTGAAGTGGAGAGCCCTGGTCTGT 675  
787 GAGGAGACCTTCCAGAGTCTCTGCTGGGGGTTTAACTCCCTGGTCTGGTCCAGCAT 846  
674 GAGGAGACCTTCCAGAGTCTCTGCTGGGGGTTTAACTCCCTGGTCTGGTCCAGCAT 615  
847 CCAGGCTGTACACCCAGATCTGCAAAATACATGTCTGGATCAATTAAGTATAGCTCC 906  
614 CCAGGCTGTACACCCAGATCTGCAAAATACATGTCTGGATCAATTAAGTATAGCTCC 555  
907 AACTGATCCAGATGCTAGCTGCTGAGCTGATCCAGATGTTATGCTCTGCTGATCCAGATG 966  
554 AACTGATCCAGATGCTAGCTGCTGAGCTGATCCAGATGTTATGCTCTGCTGATCCAGATG 495  
967 CCAGAGGCTGATGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1025  
494 CCAGAGGCTGATGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 435  
1026 GCAGTGTTCAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1085  
434 GCAGTGTTCAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375  
1086 CCCCACCTTCT 1145  
374 CCCCACCTTCT 315  
1146 GTTATTCAGAGAGAGCGAGAGCGGCTGATCCAGAGCTCTGAGAGAGTAACTGAGG 1205  
314 GTTATTCAGAGAGAGCGAGAGCGGCTGATCCAGAGCTCTGAGAGAGTAACTGAGG 255  
1206 GTACCCAGAGCTGATCT 1265  
254 GTACCCAGAGCTGATCT 195  
1266 CCGCTGTAAGCTGATCT 1335  
194 CCGCTGTAAGCTGATCT 135  
1326 GACATGTTGTGAGAGAGCTGATGATTAACATGATGATTAATCTTCAT-GTGAATGCA 1384  
134 GACATGTTGTGAGAGAGCTGATGATTAACATGATGATTAATCTTCATGATGATGCA 75  
1385 TGTAAAGGCTTAAACAGAGTGGGTGAGTGTGATTAAGGTTAACTCTCTCTCTCTCTCT 1444  
74 TGTAAAGGCTTAAACAGAGTGGGTGAGTGTGATTAAGGTTAACTCTCTCTCTCTCTCT 15  
1445 AAAAAAAAAA 1454  
14 AAAAAAAAAA 5

RESULT 3  
LOCUS BM019631 714 bp mRNA linear EST 30-OCT-2001  
DEFINITION 603647895F1 NIH\_MGC\_98 Homo sapiens cDNA clone IMAGE:5429567 5',  
RNA sequence.  
ACCESSION BM019631  
VERSION BM019631.1 GI:16533985  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
1 (bases 1 to 714)  
AUTHORS NIH-MGC <http://img.ncbi.nlm.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILLNL at:  
<http://image.illnl.gov>  
Plate: L10M1898 row: m column: 24  
High quality sequence stop: 708.

## FEATURES

Location/Qualifiers  
1..714

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5429567"  
/issue\_type="astrocytoma grade IV, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_98"  
/note="T-Organ: Brain; Vector: pORF7, Site 1: XhoI, Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adapter: GGCACGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC library."

BASE COUNT 116 a 256 c 220 g 122 t

## ORIGIN

Query Match 47.3%; Score 687.4; DB 12; Length 714;  
Best Local Similarity 99.4%; Pred. No. 2,2e-128;  
Matches 701; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

1 ACCAGGCGGAGACCAAGGCGAGGCGAGAGGACAGCTGTGGGTCCCTCCCTCTCTATC 60  
5 ACCAGGCGGAGACCAAGGCGAGGCGAGAGGACAGCTGTGGGTCCCTCCCTCTCTATC 64  
61 GCGACCTCC---CAGATCTGCGGCAAGAGAGCTGCGCACTCCACCTCCCGCGGCTCT 117  
65 GCGACCTCCAGAGAGATCTGCGGCAAGAGAGCTGCGCACTCCACCTCCCGCGGCTCT 124  
118 GCGGCGCGGCTCTGCGGAGAGCTGCGGCTGCTGATGAGCGCACTGCGGCGGAGAG 177  
125 GCGGCGCGGCTCTGCGGAGAGCTGCGGCTGCTGATGAGCGCACTGCGGCGGAGAG 184  
178 GCGGCGGCTCTGCGGAGAGAGCGGCTGCGGAGAGCGGCTGCGGAGAGCGGCTGCGG 237  
185 GCGGCGGCTCTGCGGAGAGAGCGGCTGCGGAGAGCGGCTGCGGAGAGCGGCTGCGG 244  
238 GCGGCGGCTCTGCGGAGAGAGCGGCTGCGGAGAGCGGCTGCGGAGAGCGGCTGCGG 297  
245 GCGGCGGCTCTGCGGAGAGAGCGGCTGCGGAGAGCGGCTGCGGAGAGCGGCTGCGG 304  
298 GGTGTCTCTGAG 357  
305 GGTGTCTCTGAG 364  
358 TGGGCTGAGTGGAG 417  
365 TGGGCTGAGTGGAG 424  
418 ACTGCTCTGTTGTCATCCCAAGTACACAGAGGCTCAGGCGCATCTCTGCAAGAGCA 477  
425 ACTGCTCTGTTGTCATCCCAAGTACACAGAGGCTCAGGCGCATCTCTGCAAGAGCA 484  
478 ACCGATGACAGAGATGATGTTGCTAAAGCTGCGCAGGCGCGTATGTCGGGAGCCCG 537

Db 485 AGCGATGACAGATCTCATGTTGCTAAAGCTGGCCAGGCCCCGTAGTGCAGGAGCCCCC 544  
 QY 538 GTCCGGGCGCTGAGCTTCCTACCGCTGTGCTACAGCCGAGACAGAGTGGT 597  
 Db 545 GTCCGGGCGCTGAGCTTCCTACCGCTGTGCTACAGCCGAGACAGAGTGGT 604  
 QY 598 GGTGGGGGACACAGCGGCGCCGAGAGTGAAGTACACAGAGGCTGCTGCTCAGC 657  
 Db 605 GGTGGGGGACACAGCGGCGCCGAGAGTGAAGTACACAGAGGCTGCTGCTCAGC 664  
 QY 658 ATCACTATCTGAGCCCTAAGAGTGTAGTCTTCTACCTGGC 702  
 Db 665 ATCACTATCTGAGCCCTAAGAGTGTAGTCTTCTACCTGGC 709

RESULT 4  
 BG678912 758 bp mRNA linear EST 01-MAY-2001  
 LOCUS BG678912  
 DEFINITION 602624621p1 NCI\_CGAP\_Skn4 Homo sapiens cDNA clone IMAGE:4749772 5',  
 mRNA sequence.  
 BG678912  
 ACCESSION BG678912 GI:13910309  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/.  
 AUTHORS NIH-MGC  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: James Cleaver, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLNL at:  
 http://image.llnl.gov  
 Plate: LLM10603 row: 1 column: 05  
 High quality sequence stop: 755.  
 Location/Qualifiers

FEATURES  
 source  
 1..758  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4749772"  
 /tissue\_type="squamous cell carcinoma"  
 /lab\_host="DH10B (TI phage-resistant)"  
 /clone\_id="NCI CGAP Skn4"  
 /note="Organ: Skin; Vector: pCMV-SPORT6, Site\_1: NciT,  
 Site\_2: SalI; Cloned unidirectionally. Primer: 0.150 dt.  
 Average insert size 1.5kb. Library constructed by Life  
 Technologies. Note: this is a NCI CGAP Library."  
 BASE COUNT 144 a 266 c 184 g 164 t  
 ORIGIN

Query Match 46.7%; Score 679.2; DB 10; Length 758;  
 Best Local Similarity 97.6%; Fred. No. 9.9e-127;  
 Matches 742; Conservative 0; Mismatches 13; Indels 5; Gaps 5;

QY 352 CCAGTGTGGGCTCGAGTAGGAGATATACCTGCTGCTTCAAGGGGAGACACTCGC 411  
 Db 1 CCAGTGTGGGCTCGAGTAGGAGATATACCTGCTGCTTCAAGGGGAGACACTCGC 60  
 QY 412 CGAGACTGCTGCTGTGTCCATCCAGAGTACACAGAGGCTCAGAGCCCATCTGCA 471  
 Db 61 CGAGC-ACCTGCTGTGTCCATCCAGAGTACACAGAGGCTCAGAGCCCATCTGCA 118  
 QY 472 AGGCGAAGGATGACAGATCTCATGTTGCTAAAGCTGGCCAGGCCCCGTAGTGCAGG 531

Db 119 AGGCGAAGGATGACAGATCTCATGTTGCTGAAGCTGGCCAGGCCCCGTAGTGCAGG 178  
 QY 532 CCCCCGCGCGGCGCTGAGCTTCCTACCGCTGTGCTACGCCGAGACAGTGGCAG 591  
 Db 179 CCCCCGCT-CGCGGCGCTGAGCTTCCTACCGCTGTGCTGAGCCGAGACAGTGGCAG 237  
 QY 592 GTTGTGGCTGGGACACAGCGGCGCCGAGGAGTGAAGTACACAGAGGCTGAGCTG 650  
 Db 238 GTTGTGGCTGGGACACAGCGGCGCCGAGGAGTGAAGTACACAGAGGCTGAGCTG 297  
 QY 651 CTCAGACTCATCTATCTGAGCCCTAAGAGTGTAGTCTTCTACCTGGCAGTGGTAC 710  
 Db 298 CTCAGACTCATCTATCTGAGCCCTAAGAGTGTAGTCTTCTACCTGGCAGTGGTAC 357  
 QY 711 CAACAAGTATGATGTGTGCTGACCTGAGCCGAGGACAGACCTTCCAGAGTACTGG 770  
 Db 358 CAACAAGTATGATGTGTGCTGACCTGAGCCGAGGACAGACCTTCCAGAGTACTGG 417  
 QY 771 AGGCCCCCTGCTGTGACAGAGACCTTCCAGAGGATCTCTGAGGAGTTTACCCCTG 830  
 Db 418 AGGCCCCCTGCTGTGACAGAGACCTTCCAGAGGATCTCTGAGGAGTTTACCCCTG 477  
 QY 831 TGGCTTGGCCAGATCCAGCTGTCTACACCCAGATCTGCAATATATGCTGGATCA 890  
 Db 478 TGGCTTGGCCAGATCCAGCTGTCTACACCCAGATCTGCAATATATGCTGGATCA 537  
 QY 891 TAAAGTCATAGCTCCCAATGATCCAGATGCTACGCTCCAGCTGATCCAGTGTATGCT 950  
 Db 538 TAAAGTCATAGCTCCCAATGATCCAGATGCTACGCTCCAGCTGATCCAGTGTATGCT 597  
 QY 951 CCTGCTATCCAGA-TGCCAGAGGCTCCATCTGCTCTCTCTCCAGTCCGCTG 1009  
 Db 598 CCTGCTATCCAGATCCAGATGCTCCAGAGGCTCCATCTGCTCTCTCTCCAGTCCGCTG 657  
 QY 1010 AACTCTCCCTTGTGTGCTGCTTCAAACTGTGCGCGCTCCAGACTTAAACATCTC 1069  
 Db 658 AACTCTCCCTTGTGTGCTGCTTCAAACTGTGCGCGCTCCAGACTTAAACATCTC 717  
 QY 1070 CCTCTCAGCTATCTCCCACTATCCCATCTGCTGCGC 1109  
 Db 718 CCTCTCAGCTATCTCCCACTATCCCATCTGCTGCGC 757

RESULT 5  
 BG469321 794 bp mRNA linear EST 21-MAR-2001  
 LOCUS BG469321  
 DEFINITION 602533165F1 NIH\_MGC\_15 Homo sapiens cDNA clone IMAGE:4660528 5',  
 mRNA sequence.  
 BG469321  
 ACCESSION BG469321 GI:13401596  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/.  
 AUTHORS NIH-MGC  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Sequencing by: NIH Intramural Sequencing Center  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLNL at:  
 http://image.llnl.gov  
 Plate: LLM1458 row: b column: 17  
 High quality sequence stop: 728.  
 Location/Qualifiers

FEATURES  
 source  
 1..794  
 /organism="Homo sapiens"



```

Db 301 GCTGACGCGCCGACATCCGGAACACACACTGTGGCTGAGTAGGGGATGACCACT 360
Qy 384 GCTGCTCTTTCAGGCGGAGAGCTCCGCGGAGCACTGCTTGTGTCATCCCAAGTA 443
Db 361 GCTGCTTCTTTCAGGAGAGAGCTCCGCGGAGCACTGCTTGTGTCATCCCAAGTA 420
Qy 444 CCACGAGGCTCAGGCGCCCATCTCTCCAGAGCGAATGAGTACGATCTCAATGTTCT 503
Db 421 CCACGAGGCTCAGGCGCCCATCTCTCCAGAGCGAATGAGTACGATCTCAATGTTCT 480
Qy 504 AAAGCTGGCCAGGCGCCCATCTCTCCAGAGCGAATGAGTACGATCTCAATGTTCT 563
Db 481 GAGCTGGCCAGGCGCCCATCTCTCCAGAGCGAATGAGTACGATCTCAATGTTCT 540
Qy 564 CTGTGCTCAGCGCGGAGAGCAAGTGCAGAGTCTGCTGAGGAGCAACGCGCGCGAG 623
Db 541 CTGTGCTCAGCGCGGAGAGCAAGTGCAGAGTCTGCTGAGGAGCAACGCGCGAG 599
Qy 624 AGTGAAGTACAAACAGAGGCTGAGCTGCTCCAGCATCACTATCCAGAGCCCTAAAGAGTG 683
Db 600 AGTGAAGTACAAACAGAGGCTGAGCTGCTCCAGCATCACTATCCAGAGCCCTAAAGAGTG 658
Qy 684 TGAAGCTCTTCACTCCCTGCGGCTGCTGACCAACAACTATGTTGCTGAGCTGAGCGGAG 743
Db 659 TGAAGCTCTTCACTCCCTGCGGCTGCTGACCAACAACTATGTTGCTGAGCTGAGCGGAG 717

```

```

RESULT 7
BM983288 686 bp mRNA linear EST 20-FEB-2003
LOCUS UI-CF-DUI-aav-1-20-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
DEFINITION UI-CF-DUI-aav-1-20-0-UI 3', mRNA sequence.
ACCESSION BM983288
VERSION BM983288.1 GI:19607650
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 686)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7121
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 497-634, >MIR#SINE/MIR
Seq primer: M13 FORWARD
POLYA=yes.

```

## FEATURES

## SOURCE

```

Location/Qualifiers
1..686
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-aav-1-20-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

```

```

/clone lib="UI-CF-DUI"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-DUI is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GGCTGAGGCG.
TAG LIB=UI-CF-DUI
TAG TISSUE=Lung Epithelial Cells Tissue nos 359-368
TAG_SEQ=GGCTGAGGCG"
BASE COUNT 147 a 210 c 143 g 186 t
ORIGIN

```

```

Query Match 44.9%; Score 652.8; DB 12; Length 686;
Best Local Similarity 99.6%; Pred. No. 2.1e-121;
Matches 665; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```

```

Qy 746 AGAAGCCCTGGCAGAGTACTGTGAGAGCCCTGCTGTGTGACGAGACCTCCAGAGCA 805
Db 19 AGAAGCCCTGGCAGAGTACTGTGAGAGCCCTGCTGTGTGACGAGACCTCCAGAGCA 78
Qy 806 TCTCTCTGCGGGGTGTTTACCCCTGTGCTTGTGCGGAGCATCCAGAGCTTACACCCGCA 865
Db 79 TCTCTCTGCGGGGTGTTTACCCCTGTGCTTGTGCGGAGCATCCAGAGCTTACACCCGCA 138
Qy 866 TCTGCAATACATGCTCCGATCATATAAGTCATACGCTCCAGTATCATCATGATCTTACG 925
Db 139 TCTGCAATACATGCTCCGATCATATAAGTCATACGCTCCAGTATCATCATGATCTTACG 198
Qy 926 CTCGAGCTGATCCAGATGTTATGCTCTGTGATTCAGATGCGCCAGAGCTTCATCTGTC 985
Db 199 CTCGAGCTGATCCAGATGTTATGCTCTGTGATTCAGATGCGCCAGAGCTTCATCTGTC 258
Qy 966 ATCTCTTCTCTCCAGAGTGTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1045
Db 259 ATCTCTTCTCTCCAGAGTGTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 318
Qy 1046 GCCCTCCACACTTAAACATCTCCCTCTGACCTCATCTCCCTCCACCTATCTCCATCTTC 1105
Db 319 GCCCTCCACACTTAAACATCTCCCTCTGACCTCATCTCCCTCCACCTATCTCCATCTTC 378
Qy 1106 TCCCTGTAAGTGAAGTGAATGCAAGAGTGTGCAAGGTTTATTCAGAGAGCCAG 1165
Db 379 TCCCTGTAAGTGAAGTGAATGCAAGAGTGTGCAAGGTTTATTCAGAGAGCCAG 438
Qy 1166 GAAGCGGTCATCACCCAGCCCTGTGAGAGCAAGTTACTGGGGTCAACCACTGACTTCT 1225
Db 439 GAAGCGGTCATCACCCAGCCCTGTGAGAGCAAGTTACTGGGGTCAACCACTGACTTCT 498
Qy 1226 CTGCACTCCCGCGTGTGTGACTTGGCGAAGCGTCCCTCTCTGTAACCTCACTGTTTC 1285
Db 499 CTGCACTCCCGCGTGTGTGACTTGGCGAAGCGTCCCTCTCTGTAACCTCACTGTTTC 558
Qy 1286 CTGATTCGAAATGAGAAATGACAGTCCCTTCTAGACATGTTGAGAGAGCTA 1345
Db 559 CTGATTCGAAATGAGAAATGACAGTCCCTTCTAGACATGTTGAGAGAGCTA 618
Qy 1346 TGATATACATGTGTATTAATCTTCAT-GTATTTGTCATGTAAGCTTAAACAGAGTG 1404
Db 619 TGATATACATGTGTATTAATCTTCATGATGATGATGATGATGATGATGATGATGATGAT 678
Qy 1405 GTGTGAG 1412
Db 679 GTGTGAG 686

```

us-10-021-368-2.r

[illegible]





QY 967 CCGAGAGCTCATCGTCATCCCTCTTCCCTCCAGTCGAGTCAACTCTCCCTTGTCTG 1026  
 DB 241 CCGAGAGCTCATCGTCATCCCTCTTCCCTCCAGTCGAGTCAACTCTCCCTTGTCTG 300  
 QY 1027 CACTGTTCAACCTCTGCGCGCTCCACACCTCTTAAACATCTCCCTCTCACTTATCC 1086  
 DB 301 CACTGTTCAACCTCTGCGCGCTCCACACCTCTTAAACATCTCCCTCTCACTTATCC 360  
 QY 1087 CCCACCTATCCCTCTTCTGCTCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1146  
 DB 361 CCCACCTATCCCTCTTCTGCTCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 420  
 QY 1147 TTTATTCAG 1206  
 DB 421 TTTATTCAG 480  
 QY 1207 TCACCCAACTGACTTCTCTGCACTCCCGCTGTGTGACTTGTGAGCAAGCAAGTGC 1266  
 DB 481 TCACCCAACTGACTTCTCTGCACTCCCGCTGTGTGACTTGTGAGCAAGCAAGTGC 540  
 QY 1267 CTCTCTAACCTCTGCTTCTCTCACTCTGCAAAATGGGAACAATGACCTTACTCTTAG 1326  
 DB 541 CTCTCTAACCTCTGCTTCTCTCACTCTGCAAAATGGGAACAATGACCTTACTCTTAG 600  
 QY 1327 ACATGTTGTAGAGAGACTATGATAT 1351  
 DB 601 ACATGTTGTAGAGAGACTATGATAT 625

RESULT 11  
 A1859367 736 bp mRNA linear EST 07-MAR-2000  
 LOCUS A1859367/1  
 DEFINITION un10f02.x1 NCI CGAP U4 Homo sapiens cDNA clone IMAGE:2435547 3',  
 similar to SW:FL1\_HUMAN 043240 PROTEASE SERINE-LIKE 1 PRECURSOR ;,  
 mRNA sequence.  
 A1859367.1 GI:5512983  
 EST.  
 VERSION A1859367.1  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Bukayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 736)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgsbds-remail.nih.gov  
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/HLN at:  
 www-bio1.nhl.gov/bdrip/image/image.html  
 Insert length: 913 Std Error: 0.00  
 Seq primer: -40UP from Gdbco  
 High quality sequence stop: 118.  
 Location/Qualifiers  
 1. 736  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2435547"  
 /tissue\_type="serous papillary carcinoma, high grade, 2  
 pooled tumors"  
 /lab\_host="DH10B"  
 /clone\_idb="NCI CGAP U4"  
 /note="Organ: uterus; Vector: pCMV-Sport6; Site 1: salI;  
 Site 2: NciI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.48 kb. Life Technologies catalog #:

BASE COUNT 143 a 207 c 243 g 138 t 5 others  
 ORIGIN  
 Query Match 42.6%; Score 619.2; DB 9; Length 736;  
 Best Local Similarity 92.4%; Pred. No. 1,2e-114;  
 Matches 681; Conservative 0; Mismatches 53; Indels 3; Gaps 3;  
 11542-016"  
 QY 176 AGCGGCGCTGCTCCCAAAACAGACAGCGCTTGAGACCCCGAGCTTATGCGCCCGT 235  
 DB 735 AGAGCGCGCTGCTCCCAAAACAGACAGCGCTTGAGACCCCGAGCTTATGCGCCCGT 676  
 QY 236 GCGGCGCGCGCTGCGAGCGCTTGAGAGCTTCTTCAAGCGCTCTGCTTCACTGCG 295  
 DB 675 GCGGCGCGCGCTGCGAGCGCTTGAGAGCTTCTTCAAGCGCTCTGCTTCACTGCG 617  
 QY 296 CCGGCTCTGCTGAGACAGAGTGGGTCTGAGCGCGCGAGCTGCGAGAAAGAGCAG 355  
 DB 616 CCGGCTCTGCTGAGACAGAGTGGGTCTGAGCGCGCGAGCTGCGAGAAAGAGCAG 557  
 QY 356 TGTGGCTGAGTGAAGGAGATGATCACTGCTTCTTCAAGCGAGAGCTCCCGGA 415  
 DB 556 TGTGGCTGAGTGAAGGAGATGATCACTGCTTCTTCAAGCGAGAGCTCCCGGA 497  
 QY 416 CGACTGCTCTGTTGTTCATCCCAAGTACACAGAGGCTCAGGCCCATCTGCGAAGC 475  
 DB 496 CGACTGCTCTGTTGTTCATCCCAAGTACACAGAGGCTCAGGCCCATCTGCGAAGC 438  
 QY 476 GAAAGATGAGACGATCTCATGTTGCTTAAAGTGGCCAGGCCCGTATGTCGCGGCC 535  
 DB 437 GAAAGATGAGACGATCTCATGTTGCTTAAAGTGGCCAGGCCCGTATGTCGCGGCC 378  
 QY 536 GCGTCCGCGCGCTGCGAGCTTCTTCAAGCTGCTGCTGAGAGAGAGAGAGAGAGTGG 555  
 DB 377 GCGTCCGCGCGCTGCGAGCTTCTTCAAGCTGCTGCTGAGAGAGAGAGAGAGTGG 318  
 QY 556 CTGCTGAGGAGACCAAGCGCGCGCGAGAGAGTGAAGTGAACAAGAGGCTGCTGCA 655  
 DB 317 CTGCTGAGGAGACCAAGCGCGCGCGAGAGAGTGAAGTGAACAAGAGGCTGCTGCA 258  
 QY 656 GCATCACTATCTGAGCCCTTAAAGAGTGAAGTCTTCAAGCTGCGGTGATCAACACA 715  
 DB 257 GCATCACTATCTGAGCCCTTAAAGAGTGAAGTCTTCAAGCTGCGGTGATCAACACA 198  
 QY 716 ACATGATATGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 774  
 DB 197 ACATGATATGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 138  
 QY 775 CCGCTGCTGCTGAGAGAGCCCTCAAGGATCTCTGCGGAGTGTATCCCTGAGGC 834  
 DB 137 CCGCTGCTGCTGAGAGAGCCCTCAAGGATCTCTGCGGAGTGTATCCCTGAGGC 78  
 QY 835 TCTGCCAGAGATCAGCTGCTTACACCCAGATCTGCAAAATCATGTCCTGATCAATAA 894  
 DB 77 TCTGCCAGAGATCAGCTGCTTACACCCAGATCTGCAAAATCATGTCCTGATCAATAA 18  
 QY 895 GTCATACGCTCCCAACTG 911  
 DB 17 GTCATACGCTCCCAACTG 1

RESULT 12  
 BE899455 866 bp mRNA linear EST 29-SEP-2000  
 LOCUS BE899455  
 DEFINITION 60162224F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3952448 5',  
 mRNA sequence.  
 BE899455  
 BE899455.1 GI:10366783  
 EST.  
 VERSION BE899455  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Bukayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 866)  
 AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: DCTD/DP  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LHC822 row: c column: 09  
 High quality sequence start: 5  
 High quality sequence stop: 692.  
 Location/Qualifiers  
 1..866  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3952448"  
 /tissue\_type="adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_9"  
 /note="Organ: ovary; Vector: pOT7, site\_1: XhoI, site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 157 a 282 c 261 g 166 t  
 ORIGIN  
 Query Match 41.3%; Score 601; DB 10; Length 866;  
 Best Local Similarity 93.5%; Pred. No. 5.7e-111;  
 Matches 706; Conservative 0; Mismatches 35; Indels 14; Gaps 7;  
 QY 22 GGGGAGAGGACGCTGGGGTCCCTCCCTCCCTCCATGGGAGCTCCGAGTCCCTGAGC 81  
 DB 1 GGGGAGAGGACGCTGGGGTCCCTCCCTCCCTCCATGGGAGCTCCGAGTCCCTGAGC 60  
 QY 82 ATGAGAGCTCGGACCTCCACCTCTCCGCGGCTCTGGGCGCCGGGCTCTGGCGAAGCTG 141  
 DB 61 ATGAGAGCTCGGACCTCCACCTCTCCGCGGCTCTGGGCGCCGGGCTCTGGCGAAGCTG 120  
 QY 142 CTGGCGGCTGATAGGCGCACTCTGGGCGGAGAGGCGGCGCTCTCCGCAAAAGCAGC 201  
 DB 121 CTGGCGGCTGATAGGCGCACTCTGGGCGGAGAGGCGGCGCTCTCCGCAAAAGCAGC 180  
 QY 202 AGCGGCTTGAACCCGGAAGCTATAGGCGCGCGTGCAGGCGGCTCGAGCCTTGAGGAG 261  
 DB 181 AGCGGCTTGAACCCGGAAGCTATAGGCGCGCGTGCAGGCGGCGCTCGAGCCTTGAGGAG 240  
 QY 262 GTCTGCTTTCAAGGCGCTCTCGTTCCACTGCGGGGTGTCTGTGGAGCCAGAGTTGG 321  
 DB 241 GTCTGCTTTCAAGGCGCTCTCGTTCCACTGCGGGGTGTCTGTGGAGCCAGAGTTGG 300  
 QY 322 GTGGGAGGCGGCGGCACTGGGAAACAAGCACTGTGGGCTGAGTGGGGAGATATAC 381  
 DB 301 GTGGGAGGCGGCGGCACTGGGAAACAAGCACTGTGGGCTGAGTGGGGAGATATAC 360  
 QY 382 GTGTGCTTTCAAGGCGGAGAGCTCCGCGGAGCACTCGCTGTGTGTCCATCCAG 441  
 DB 361 GTGTGCTTTCAAGGCGGAGAGAG-TCCGCGGAGCACTCGCTGTGTGTCCATCCAG 419  
 QY 442 TACGACGAGGCTTCAAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 501  
 DB 420 TACGACGAGGCTTCAAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 479  
 QY 502 CTAAAGCTGAGGCGGAG 561

DB 480 CTGAGAGCTGAC--AGCGCTAGTGTGCGGAGCCCGGCT--CGGGCCCTGCGAG--TTCCCTA 534  
 QY 562 CGCTGTCTCAGACCCCGGAGAGACAGTGTGCTGTGGCTGGGCGACACGCGCCCGG 621  
 DB 535 CCGTGTCTCAGACCCCGGAGAGACAGTGTGCTGTGGCTGGGCGACACGCGCCCGG 590  
 QY 622 AGAGTGAAGTACAACAAGAGGCTGACCTGCTCCAGCATCATCTGAGCCCTTAAGAG 681  
 DB 591 AGAGTGAAGTACAACAAGAGGCTGACCTGCTCCAGCATCATCTGAGCCCTTAAGAG 647  
 QY 682 TGTGAGGCTTTTACCTGTGGCGGTGTGACCAACAACATGATGTGTGACTGGACCGG 741  
 DB 648 TGTGAGGCTTTTACCTGTGGCGGTGTGACCAACAACATGATGTGTGACTGGACCGG 707  
 QY 742 GGCAGAGACCTTGGCGAGTGTGACTGTGAGGCGCC 776  
 DB 708 GG-CAGGACCTTGGCGAGTGTGAGTGTGAGGCGCC 741  
 RESULT 13  
 CB216785  
 LOCUS 614 bp mRNA linear EST 06-FEB-2003  
 DEFINITION NISC\_nq08h06.y1 NICH\_H5\_Uc2 Homo sapiens CDNA clone IMAGE:5938427  
 ACCESSION CB216785  
 VERSION CB216785.1 GI:28264977  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 614)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 CDNA Library Preparation: CDNA Library Arrayed by: The I.M.A.G.E. Consortium/ULNL  
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LHM3166 row: P column: 12  
 Seq primer: M13RP1 reverse primer (ABI).  
 Location/Qualifiers  
 1..614  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5938427"  
 /sex="female"  
 /tissue\_type="normal endometrium, mid-secretory phase, cycle day 23"  
 /lab\_host="DH10B (T1-resistant)"  
 /clone\_lib="NICH\_H5\_Uc2"  
 /note="Organ: uterus; Vector: pCMV-SPORT6.1.cdb (ResGen, Invitrogen Corporation); Site\_1: NotI; Site\_2: EcoRV; Cloned unidirectionally from macroquantity amounts of mRNA from normal endometrial tissue (mid-secretory phase, cycle day 23). Average insert size 1.6 kb. Library constructed by ResGen (Invitrogen Corporation)."  
 BASE COUNT 94 a 225 c 190 g 105 t  
 ORIGIN  
 Query Match 41.3%; Score 600.4; DB 14; Length 614;  
 Best Local Similarity 99.0%; Pred. No. 7.5e-111;  
 Matches 604; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 ACCAGCGGAGACGACGAGGCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60



DEFINITION wml10h02.x1 NCI CGAP Ut4 Homo sapiens cDNA clone IMAGE:2435571 3' similar to SW:PSL1\_HUMAN 043240 PROTEASE SERINE-LIKE 1 PRECURSOR ; mRNA sequence.

ACCESSION A1859383

VERSION A1859383.1 GI:5512999

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapsb-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.lnl.gov/bbrp/image/image.html  
Insert Length: 919 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 423.

FEATURES  
source  
1..727  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2435571"  
/tissue\_type="serous papillary carcinoma, high grade, 2 pooled tumors"  
/lab\_host="DH10B"  
/clone\_1ib="NCI CGAP Ut4"  
/note="Organ: uterus; Vector: pCMV-Sport6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.48 kb. Life Technologies catalog #: 11542-016"

BASE COUNT 137 a 208 c 235 g 132 t 15 others

ORIGIN

Query Match 39.9%; Score 580; DB 9; Length 727;  
Best Local Similarity 91.5%; Pred. No. 9.7e-107;  
Matches 671; Conservative 0; Mismatches 55; Indels 7; Gaps 6;

QY 181 GCGTGCTCCCCCAAAACGACACGCGCTTGACCCCGAAGCTTATGGCGCCCGTGGCG 240  
DB 727 GCGTGCTCCCCCAAAACGACACGCGCTTGACCCCGAAGCTTATGGCGCCCGTGGCG 668

QY 241 GCGGCGTGCAGCCCTGGGAGGCTCGCTCTTCAACGCGGCTCTGCTCACTGCGGAGT 300  
DB 667 GCG-GCTMGACGCTGGGAGGCTCGCTCTTCAACGCGGCTCTGCTCACTGCGGAGT 611

QY 301 GTCTGTGTGACACAGAGTTGGTGTGACGCGCGCGCACTGGGAAACAGCAAGCACTGTGG 360  
DB 610 TGTCTGTGTGACACAGAGTTGGTGTGACGCGCGCGCACTGGGAAACAGCAAGCACTGTGG 552

QY 361 GGTTCAGTGGGAGATATCACTGTGCTTTTCAAGGGGAGACACTCGCGGAGCACT 420  
DB 551 GGTTCAGTGGGAGATATCACTGTGCTTTTCAAGGGGAGACACTCGCGGAGCACT 494

QY 421 CGCTCTGTGTCCATCCCAAGTACACACAGGCTCAGGCCCATCTGCGCAAGGCAAG 480  
DB 493 CGCTCTGTGTCCATCCCAAGTACACACAGGCTCAGGCCCATCTGCGCAAGGCAAG 434

QY 481 GATGAGCAGATCTCATGTGCTTAAAGCTGCGCAGGCGCCGCTAGTCCGCGGCGTCC 540  
DB 433 GATGAGCAGATCTCATGTGCTTAAAGCTGCGCAGGCGCCGCTAGTCCGCGGCGTCC 374

QY 541 CGGCGCTGAGCTTCCCTACCGCTGTGCTCAGCCCGGAGACCAAGTCCAGAGTTGCTGCG 600  
DB 373 CGGCGCTGAGCTTCCCTACCGCTGTGCTCAGCCCGGAGACCAAGTCCAGAGTTGCTGCG 314

QY 601 TGGGAGCACACGCGCCCGGAGAGTGAATACACAGGCGCTGACCTGCTCCAGCATC 660  
DB 313 TGGGAGCACACGCGCCCGGAGAGTGAATACACAGGCGCTGACCTGCTCCAGCATC 254

QY 661 ACTATCTGAGCCCTTAAAGTGTGAGTCTTCTACCTGCGCGGTGACCAACCAATG 720  
DB 253 ACTATCTGAGCCCTTAAAGTGTGAGTCTTCTACCTGCGCGGTGACCAACCAATG 194

QY 721 ATATGCTGAGCTGAGACG-GGGGCGAGACCTTGCCAGAGTACTCTGAGGCGCCCT 779  
DB 193 ATATGCTGAGCTGAGACGCGGGGCGAGACCTTGCCAGAGTACTCTGAGGCGCCCT 134

QY 780 GGTCTGTGACGAGACCTTCCAGGAGTCTGCTGTGGGGTGTTCACCCCTGTGCTTGC 839  
DB 133 GGTCTGTGACGAGACCTTCCAGGAGTCTGCTGTGGGGTGTTCACCCCTGTGCTTGC 74

QY 840 CCAAGCATCCAGCTGTGTACACCCAGATCTGCAATATACATGCTGATCAATTAAGTCAT 899  
DB 73 CCAAGCATCCAGCTGTGTACACCCAGATCTGCAATATACATGCTGATCAATTAAGTCAT 14

QY 900 ACGCTCCACTGA 912  
DB 13 ACGCTCCACTGA 1

Search completed: November 25, 2003, 07:29:52  
Job time : 3311 secs

\_\_\_\_\_

```

Db 121 GCCCGGCTCTGGCGAAGCTGCTGCGCTGCTGATGCGCGCAACTCTGAGGCGCGAGAGGCG 180
QY 181 GCGCTGCTCCCCAAAACGACACGCGCTTGAGACCCCGAAGCTATGCGCCCTGCTGCGCG 240
Db 181 GCGCTGCTCCCCAAAACGACACGCGCTTGAGACCCCGAAGCTATGCGCCCTGCTGCGCG 240
QY 241 CCGGAGCTCCGAGCCCTGGAGGCTGCTGCTTCAAGGCGCTGCTGCTGCTGCTGCTGCTG 300
Db 241 CCGGAGCTCCGAGCCCTGGAGGCTGCTGCTTCAAGGCGCTGCTGCTGCTGCTGCTGCTG 300
QY 301 GTCCTGATGAGCAGAGATGAGTGTGCTGACGCGCGCGCACTGCGGAAACAGCCACTGTG 360
Db 301 GTCCTGATGAGCAGAGATGAGTGTGCTGACGCGCGCGCACTGCGGAAACAGCCACTGTG 360
QY 361 GCTCGAGTAGGAGGATGATCACTGCTGCTGCTTCTTCAAGGCGCGAGCACTGCGCGAG 420
Db 361 GCTCGAGTAGGAGGATGATCACTGCTGCTGCTTCTTCAAGGCGCGAGCACTGCGCGAG 420
QY 421 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
Db 421 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
QY 481 GATGAGCAGATCTCATGTTGCTAAAGCTGCGCAAGCGCGCTGCTGCTGCTGCTGCTG 540
Db 481 GATGAGCAGATCTCATGTTGCTAAAGCTGCGCAAGCGCGCTGCTGCTGCTGCTGCTG 540
QY 541 CCGGCGCTGAGAGCTTCCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 541 CCGGCGCTGAGAGCTTCCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
QY 601 TGGGCGCACACGCGCGCGCGAGAGTAGAACAACAGGCGCTGACCTGCTGCAAGATC 660
Db 601 TGGGCGCACACGCGCGCGCGAGAGTAGAACAACAGGCGCTGACCTGCTGCAAGATC 660
QY 661 ACTATCTGAGCCCTTAAAGAGTGTGAGGCTTCTTCAACCTGCGCTGCTGCTGCTG 720
Db 661 ACTATCTGAGCCCTTAAAGAGTGTGAGGCTTCTTCAACCTGCGCTGCTGCTGCTG 720
QY 721 ATATGCTGAGCTGAGACCGGCGCGAGACCCCTTGCAGAGTAGACTCTGAGAGCGCC 780
Db 721 ATATGCTGAGCTGAGACCGGCGCGAGACCCCTTGCAGAGTAGACTCTGAGAGCGCC 780
QY 781 GTCTGTGAGAGACCCCTTCAAGGATCTCTGCTGCGGCTGTTTACCCCTGCTGCTG 840
Db 781 GTCTGTGAGAGACCCCTTCAAGGATCTCTGCTGCGGCTGTTTACCCCTGCTGCTG 840
QY 841 CAGATCTCAGCTGCTTACACCCAGATCTGCAATACATGCTGCTGCTGCTGCTGCTG 900
Db 841 CAGATCTCAGCTGCTTACACCCAGATCTGCAATACATGCTGCTGCTGCTGCTGCTG 900
QY 901 CCGTCCAACTGATCCAGATGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 901 CCGTCCAACTGATCCAGATGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
QY 961 CAGATGCGCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db 961 CAGATGCGCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
QY 1021 TGTCTGACTGTGTTAAACCTTGTGCGCGCCTTCCACACCTCTTAAAGATCTCCCTCT 1080
Db 1021 TGTCTGACTGTGTTAAACCTTGTGCGCGCCTTCCACACCTCTTAAAGATCTCCCTCT 1080
QY 1081 CATTCCCGCACCTATCCCAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 1081 CATTCCCGCACCTATCCCAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
QY 1141 CAAAGGTTATTTCCAGAGAGCGAGAGCGGATATACCCAGCGCTTGAAGAGATTA 1200
Db 1141 CAAAGGTTATTTCCAGAGAGCGAGAGCGGATATACCCAGCGCTTGAAGAGATTA 1200
QY 1201 CTGGGGTACCCCAACTGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db 1201 CTGGGGTACCCCAACTGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260

```

```

QY 1261 AGTGCCCTCTCTGAACTCATGTTCTCTCATCTGCAAAATGSGAAATGACGTGCTAAC 1320
Db 1261 AGTGCCCTCTCTGAACTCATGTTCTCTCATCTGCAAAATGSGAAATGACGTGCTAAC 1320
QY 1321 TCTTAGACATGTTGTGAGAGACATGATATATACATGCTGATGTAAATCTTCAATGAT 1380
Db 1321 TCTTAGACATGTTGTGAGAGACATGATATATACATGCTGATGTAAATCTTCAATGAT 1380
QY 1381 GTCAATGTAAGCTTATACACAGTGGGTGCTGAGTTCTGACTAAAGSTACTGTTGCTG 1440
Db 1381 GTCAATGTAAGCTTATACACAGTGGGTGCTGAGTTCTGACTAAAGSTACTGTTGCTG 1440
QY 1441 AAAAAAAAAAAAAA 1454
Db 1441 AAAAAAAAAAAAAA 1454

RESULT 2
US-10-301-822-94
Sequence 94, Application US/10301822
Publication No. US20030148410A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Berger, Allison
APPLICANT: Guillemette, Tracy L.
APPLICANT: Kametkar, Shubhangi
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Thibodeau, Stephen N.
APPLICANT: Burgart, Lawrence J.
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF COLON CANCER
FILE REFERENCE: MP001-029P2RM
CURRENT APPLICATION NUMBER: US/10/301,822
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US 60/339,971
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 60/361,978
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/381,988
PRIOR FILING DATE: 2002-05-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 94
LENGTH: 1454
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (82)...(912)
US-10-301-822-94

Query Match 100.0%; Score 1454; DB 12; Length 1454;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1454; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 241 CGGAGCTGAGAGCCCTGAGAGCTGCTCTTCAACGAGCTCTGCTTCCACTGCGCGAGT 300  
 DB 241 CGGAGCTGAGAGCCCTGAGAGCTGCTCTTCAACGAGCTCTGCTTCCACTGCGCGAGT 300  
 QY 301 GTCCTGAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360  
 DB 301 GTCCTGAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360  
 QY 361 GCTGAGAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 420  
 DB 361 GCTGAGAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 420  
 QY 421 GCTGAGAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 480  
 DB 421 GCTGAGAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 480  
 QY 481 GATGAGAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 540  
 DB 481 GATGAGAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 540  
 QY 541 CGGAGCTGAGAGCCCTGAGAGCTGCTCTTCAACGAGCTCTGCTTCCACTGCGCGAGT 600  
 DB 541 CGGAGCTGAGAGCCCTGAGAGCTGCTCTTCAACGAGCTCTGCTTCCACTGCGCGAGT 600  
 QY 601 TGGGAG 660  
 DB 601 TGGGAG 660  
 QY 661 ACTATGCTGAG 720  
 DB 661 ACTATGCTGAG 720  
 QY 721 ATATGCTGAG 780  
 DB 721 ATATGCTGAG 780  
 QY 781 GTCCTGAG 840  
 DB 781 GTCCTGAG 840  
 QY 841 CAGATCAG 900  
 DB 841 CAGATCAG 900  
 QY 901 CGCTCAG 960  
 DB 901 CGCTCAG 960  
 QY 961 CAGATCAG 1020  
 DB 961 CAGATCAG 1020  
 QY 1021 TGTCTGAG 1080  
 DB 1021 TGTCTGAG 1080  
 QY 1081 CATTCGAG 1140  
 DB 1081 CATTCGAG 1140  
 QY 1141 CAAAGGTTTATTCAG 1200  
 DB 1141 CAAAGGTTTATTCAG 1200  
 QY 1201 CTGGGAGTCAAG 1260  
 DB 1201 CTGGGAGTCAAG 1260  
 QY 1261 AGTGCCTCTCTAAG 1320  
 DB 1261 AGTGCCTCTCTAAG 1320

QY 1321 TCTTGAAG 1380  
 DB 1321 TCTTGAAG 1380  
 QY 1381 GTCATGTAAG 1440  
 DB 1381 GTCATGTAAG 1440  
 QY 1441 AAAAAAAAAAAAAA 1454  
 DB 1441 AAAAAAAAAAAAAA 1454

## RESULT 3

US-10-366-288-51  
 ; Sequence 51, Application US/10366288  
 ; Publication No. US20030216288A1

## GENERAL INFORMATION:

; APPLICANT: Powell, Douglas  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING  
 ; TITLE OF INVENTION: AIDS AND HIV-RELATED DISORDERS USING 1414, 1481, 1553,  
 ; TITLE OF INVENTION: 34021, 1720, 1683, 1552, 1682, 1675, 12825, 9952, 5816,  
 ; TITLE OF INVENTION: 10002, 1611, 1371, 14324, 126, 270, 312, 167, 326, 18926,  
 ; FILE REFERENCE: MP102-025P1BONOMIM  
 ; CURRENT APPLICATION NUMBER: US/10/366,288  
 ; PRIOR FILING DATE: 2003-02-13  
 ; PRIOR APPLICATION NUMBER: 60/357,391  
 ; PRIOR FILING DATE: 2002-02-15  
 ; PRIOR APPLICATION NUMBER: 60/380,249  
 ; PRIOR FILING DATE: 2002-05-13  
 ; PRIOR APPLICATION NUMBER: 60/391,306  
 ; PRIOR FILING DATE: 2002-06-25  
 ; PRIOR APPLICATION NUMBER: 60/406,297  
 ; PRIOR FILING DATE: 2002-08-27  
 ; PRIOR APPLICATION NUMBER: 60/412,007  
 ; PRIOR FILING DATE: 2002-09-19  
 ; PRIOR APPLICATION NUMBER: 60/417,508  
 ; PRIOR FILING DATE: 2002-10-10  
 ; PRIOR APPLICATION NUMBER: 60/432,318  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 51  
 ; LENGTH: 1454  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-366-288-51

Query Match 100.0%; Score 1454; DB 12; Length 1454;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1454; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCAGGAG 60  
 DB 1 ACCAGGAG 60  
 QY 61 GGGAGCTCCAG 120  
 DB 61 GGGAGCTCCAG 120  
 QY 121 GGGAGCTCCAG 180  
 DB 121 GGGAGCTCCAG 180  
 QY 181 GGGAGCTCCAG 240  
 DB 181 GGGAGCTCCAG 240  
 QY 241 GGGAGCTCCAG 300  
 DB 241 GGGAGCTCCAG 300





Dp	181	GGGCTGTCCTCCCAAGACACAGCGGCTTGGACCCGGAAGCCTAATGGCCCTCCCTGGCGG	240
Qy	241	CGGCGCTGCAACCCCTGGCAAGTCTGCTCTTCAACGGCTCTCGTTCCACTGTCGGGAGT	300
Dp	241	CGGCGCTGCAACCCCTGGCAAGTCTGCTCTTCAACGGCTCTCGTTCCACTGTCGGGAGT	300
Qy	301	GTCCTGATGAGACAGAGTTGGGGTGGTGGACGGCGGCACTGTGGGAAACAAGGCACTGTGG	360
Dp	301	GTCCTGATGAGACAGAGTTGGGGTGGTGGACGGCGGCACTGTGGGAAACAAGGCACTGTGG	360
Qy	361	GCTGAGATGAGGGGATGATCACTGCTGCTTCTTCAGGGGAGACAGTCCGCGGAGACT	420
Dp	361	GCTGAGATGAGGGGATGATCACTGCTGCTTCTTCAGGGGAGAGCAGTCCGCGGAGACT	420
Qy	421	CGCTCTGTTGTCATCCCAAGAACACACAGAGGCTCAAGGCCCATCTCTGCAAGGCGGAGCG	480
Dp	421	CGCTCTGTTGTCATCCCAAGAACACACAGAGGCTCAAGGCCCATCTCTGCAAGGCGGAGCG	480
Qy	481	GATGAGACAGATCTCATGTTGTCTTAAGCTGAGCAGGCGCGATGATCCGGGAGCCCGGCGTC	540
Dp	481	GATGAGACAGATCTCATGTTGTCTTAAGCTGAGCAGGCGCGATGATCCGGGAGCCCGGCGTC	540
Qy	541	CGGGCCCTGCAAGTTCCTCACTGCTGTGCTCAGGCCCGAGACCAAGTCCAGGTTGCTGGC	600
Dp	541	CGGGCCCTGCAAGTTCCTCACTGCTGTGCTCAGGCCCGAGACCAAGTCCAGGTTGCTGGC	600
Qy	601	TGGGACACACAGCGCGCGCCGAGAGATGAAGTACACAAAGGCGCTGACCGCTCCAGCAATC	660
Dp	601	TGGGACACACAGCGCGCGCCGAGAGATGAAGTACACAAAGGCGCTGACCGCTCCAGCAATC	660
Qy	661	ACTATCCCTGAGCGCCTTAAGAGTGTGAGGTCTCTTACCCGTGGCTGATCAACAACAATG	720
Dp	661	ACTATCCCTGAGCGCCTTAAGAGTGTGAGGTCTCTTACCCGTGGCTGATCAACAACAATG	720
Qy	721	ATATGTGCTGGACCTGAGACCGGGGGGACAGGACCCCTTCCAGAGGATCTGGAGAGCCCCCTG	780
Dp	721	ATATGTGCTGAGCTGAGACCGGGGGGACAGGACCCCTTCCAGAGGATCTGGAGAGCCCCCTG	780
Qy	781	GTCCTGACAGAGACCCCTCAAGGCAATCTCTGTGGGGTGTTCACCCCTGTGGCTCTGCC	840
Dp	781	GTCCTGACAGAGACCCCTCAAGGCAATCTCTGTGGGGTGTTCACCCCTGTGGCTCTGCC	840
Qy	841	CAGATCCAGCTGTCTTACACCCAGATCTGGCAAAATCAGTCTCTGATCAATTAAGTCATTA	900
Dp	841	CAGATCCAGCTGTCTTACACCCAGATCTGGCAAAATCAGTCTCTGATCAATTAAGTCATTA	900
Qy	901	CGCTCCACTGATCAGATGTGTAAGCTCCAGCTGATGATCAGATGTATGCTCTGTGATATC	960
Dp	901	CGCTCCACTGATCAGATGTGTAAGCTCCAGCTGATGATCAGATGTATGCTCTGTGATATC	960
Qy	961	CAGATGCCCAAGAGGTTCATGTGCATCTCTCTTCTCCAGTGGGCTGAACTCTCCCT	1020
Dp	961	CAGATGCCCAAGAGGTTCATGTGCATCTCTCTTCTCCAGTGGGCTGAACTCTCCCT	1020
Qy	1021	TGTCGCACTGTTCAAACTCTGCGCGGCTCCACACCTCTAAACATCTCCCTCTCACT	1080
Dp	1021	TGTCGCACTGTTCAAACTCTGCGCGGCTCCACACCTCTCTAAACATCTCCCTCTCACT	1080
Qy	1081	CATTCGCCCAACCTATCCCATTTCTCTGCTGTATCTGAAGCTGAATGACAGAAAGTGTGG	1140
Dp	1081	CATTCGCCCAACCTATCCCATTTCTCTGCTGTATCTGAAGCTGAATGACAGAAAGTGTGG	1140
Qy	1141	CAAAAGTTATTCAGAGAGCGAGAAAGCGGATATACCAAGCGCTGTGAGAGATTA	1200
Dp	1141	CAAAAGTTATTCAGAGAGCGAGAAAGCGGATATACCAAGCGCTGTGAGAGATTA	1200
Qy	1201	CTGGAGTCAACCAACTGACTTCTCTGCTCACTCCCGCTGTGTACTTTGGGCAAGCA	1260
Dp	1201	CTGGAGTCAACCAACTGACTTCTCTGCTCACTCCCGCTGTGTACTTTGGGCAAGCA	1260
Qy	1261	AGTGCCTCTCTGAACTCAAGTTCCTCATCTGCAAAATGGGAACAATGACGTGCTTAC	1320
Dp	1261	AGTGCCTCTCTGAACTCAAGTTCCTCATCTGCAAAATGGGAACAATGACGTGCTTAC	1320

```

QY      1321  TCTTAGACATCTGTGTGGAGACATATATATTAACATGTGTATGTAATCTTCATGTGATT 1380
Db      1321  TCTTAGACATCTGTGTGGAGACATATATATTAACATGTGTATGTAATCTTCATGTGATT 1380
QY      1381  GTCATGTATAGGCTTAACACAGTGGGTGTGACTCTCACTAAAGGTAACTCGTGTGTGTG 1440
Db      1381  GTCATGTATAGGCTTAACACAGTGGGTGTGACTCTCACTAAAGGTAACTCGTGTGTGTG 1440
QY      1441  AAAAAAAAAAAAAA 1454
Db      1441  AAAAAAAAAAAAAA 1454

```

RESULT 5  
US-10-097-340-166

```

; Sequence 166, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:

```

APPLICANT: John MONAHAN  
APPLICANT: Manjula GANNAVAPATI  
APPLICANT: Sebastian HOERSCH

APPLICANT: Shubhangi KAMATKAR  
APPLICANT: Steve G. KOVATS  
APPLICANT: Rachel E. MEYERS

APPLICANT: Michael MORRISEY  
APPLICANT: Peter OLANDT  
APPLICANT: Ami SRN

APPLICANT: Peter VEIBY  
APPLICANT: Gordon B. MILLS  
APPLICANT: Robert C. DICK

APPLICANT: ROBERT C. BAST, JR.  
APPLICANT: Karen LU  
APPLICANT: Rosemarie SCHMANDT

; APPLICANT: Xumei ZHAO  
 ; APPLICANT: Karen GLATT  
 ; TITLE OF INVENTION: Nucleic Acid

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

; CURRENT FILING DATE: 2002-03-14  
 ; PRIOR APPLICATION NUMBER: 60/276,025  
 ; PRIOR FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: 60/325,149  
PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: 60/276,028  
 ; PRIOR FILING DATE: 2001-03-14  
 ; PRIOR APPLICATION NUMBER: 60/324,967

; PRIOR FILING DATE: 2001/09/26  
 ; PRIOR APPLICATION NUMBER: 60/311,732  
 ; PRIOR FILING DATE: 2001-08-10

;; PRIOR APPLICATION NUMBER: 60/325,102  
;; PRIOR FILING DATE: 2001-09-26  
;; PRIOR APPLICATION NUMBER: 60/323,580

```

; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 1.0.0

```

```

: SEQ ID NO 166
:
: LENGTH: 1454
:
: TYPE: DNA

```

ORGANISM: Homo sapiens  
US-10-097-340-166

Query Match	100.0%;	Score 1454;	DB 14;	Length 1454;
Best Local Similarity	100.0%;	Pred. No. 0;		

Matches	1454;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	ACCACGGCGACACCA	CGACGGCGACAGG	CGACGTCGTGGGTC	CCCCCTCCTCTTCT	ATC	60		

Db  
1 ACCACCGGCGACCCACAGGCGAGGGCGAGGCGACGCTGGGTCCCTCCCTTCCTTCTATC 60

**QY** 61 GGGCACTCCAGATCTGTGCGCATGAGAGCTCCGCACCTTCCACTCTCCGCGCCTCTGSC 120  
**Dp** 61 GGGCACTCCAGATCTGTGCGCATGAGAGCTCCGCACCTTCCACTCTCCGCGCCTCTGSC 120

QY 121 GCCCGGAGCTCTGGAGAGCTGCTGCGCTGCTGATGAGCGCAACTGCTGGCGCGAGAGGCG 180  
DB 121 GCCCGGAGCTCTGGAGAGCTGCTGCGCTGCTGATGAGCGCAACTGCTGGCGCGAGAGGCG 180  
QY 181 GCGCTGCTCCCCAAAACGACACGCGCTTGAACCCCGAAGCCTATGCGCGCCCGTGGCGG 240  
DB 181 GCGCTGCTCCCCAAAACGACACGCGCTTGAACCCCGAAGCCTATGCGCGCCCGTGGCGG 240  
QY 241 CCGGCGTCCAGAGCCCTGGAGAGCTGCTGCTTCAACGCGCTCTCCCTCACTGCGCGGCT 300  
DB 241 CCGGCGTCCAGAGCCCTGGAGAGCTGCTGCTTCAACGCGCTCTCCCTCACTGCGCGGCT 300  
QY 301 GTCCTGCTGAGACAGAGTGGTGGTCTGACGCGCGCGCACTGGGAAACAGCACTGTGG 360  
DB 301 GTCCTGCTGAGACAGAGTGGTGGTCTGACGCGCGCGCACTGGGAAACAGCACTGTGG 360  
QY 361 GCTCGAGTGGGGATGATCACTGCTGCTTCAAGGGGAGAGCTGCGCGGAGACT 420  
DB 361 GCTCGAGTGGGGATGATCACTGCTGCTTCAAGGGGAGAGCTGCGCGGAGACT 420  
QY 421 CCGCTGCTTGCATCCCAAGTACACACAGGCGTCAAGGCGCATCCCTGCGAAGCGAAGCG 480  
DB 421 CCGCTGCTTGCATCCCAAGTACACACAGGCGTCAAGGCGCATCCCTGCGAAGCGAAGCG 480  
QY 481 GATGAGCAGATCTCATGTTGCTAAAGCTGCGCAGGCGCGTGAAGTGGCGCGCGCGCTG 540  
DB 481 GATGAGCAGATCTCATGTTGCTAAAGCTGCGCAGGCGCGTGAAGTGGCGCGCGCGCTG 540  
QY 541 CCGGCGCTGAGCTTCCCTCAACGCGCTGCTGACGCGCGGAGACCAAGTGGCGCGCTG 600  
DB 541 CCGGCGCTGAGCTTCCCTCAACGCGCTGCTGACGCGCGGAGACCAAGTGGCGCGCTG 600  
QY 601 TGGGCGACACCGCGCGCGCGAGAGTGAAGTCAACAAAGGCGCTGACCTGCTCCAGACT 660  
DB 601 TGGGCGACACCGCGCGCGCGAGAGTGAAGTCAACAAAGGCGCTGACCTGCTCCAGACT 660  
QY 661 ACTATCTGAGCCCTTAAAGAGTGTGAGGCTTCTTCACTGCGCGTGTCAACAAAGACT 720  
DB 661 ACTATCTGAGCCCTTAAAGAGTGTGAGGCTTCTTCACTGCGCGTGTCAACAAAGACT 720  
QY 721 ATATGCTGAGCTGAGCGGCGCGCGAGCCCTTGCACAGAGTCACTGAGAGCGCGCGCTG 780  
DB 721 ATATGCTGAGCTGAGCGGCGCGCGAGCCCTTGCACAGAGTCACTGAGAGCGCGCGCTG 780  
QY 781 GTCCTGAGAGAGCCCTTCAAGGCACTCTGCTGGGAGTGTAAACCTGCTGCTGCG 840  
DB 781 GTCCTGAGAGAGCCCTTCAAGGCACTCTGCTGGGAGTGTAAACCTGCTGCTGCG 840  
QY 841 CAGCATCTGAGCTGTCTACACCCAGATCTGCAATATCATGTCTGATCATATAAGTCTA 900  
DB 841 CAGCATCTGAGCTGTCTACACCCAGATCTGCAATATCATGTCTGATCATATAAGTCTA 900  
QY 901 CCGTCAATGATCCAGATGCTACGCTCAGCTGATCCAGATGATGATGCTGCTGCTGATC 960  
DB 901 CCGTCAATGATCCAGATGCTACGCTCAGCTGATCCAGATGATGATGCTGCTGCTGATC 960  
QY 961 CAGATGCTCAGAGGCTCACTGCTCACTCTTCTCCCACTGCGCTGAGTCTGCTGCT 1020  
DB 961 CAGATGCTCAGAGGCTCACTGCTCACTCTTCTCCCACTGCGCTGAGTCTGCTGCT 1020  
QY 1021 TGTCTGCACTGTTCAAACTCTGCGCGCTTCAACAGCTTAAACATCTCCCTGCACT 1080  
DB 1021 TGTCTGCACTGTTCAAACTCTGCGCGCTTCAACAGCTTAAACATCTCCCTGCACT 1080  
QY 1081 CATTCGCCCACTATCCCACTCTGCTGCTGATGAGTGAAGTGAAGTGAAGTGAAGTGA 1140  
DB 1081 CATTCGCCCACTATCCCACTCTGCTGCTGATGAGTGAAGTGAAGTGAAGTGAAGTGA 1140  
QY 1141 CAAAGGTTTATTCAGAGAGGCGAGAGAGCGCTGATCAACAGCTCTGAGAGAGTTA 1200  
DB 1141 CAAAGGTTTATTCAGAGAGGCGAGAGAGCGCTGATCAACAGCTCTGAGAGAGTTA 1200

QY 1201 CTGGGGTACCCCAACTGACTTCTCTGCACTCCCGCTGTGACTTTGGGCAAGCCA 1260  
DB 1201 CTGGGGTACCCCAACTGACTTCTCTGCACTCCCGCTGTGACTTTGGGCAAGCCA 1260  
QY 1261 AGTGGCTCTCTGAACTCACTTCTCTGATCTGCAAAATGGGAAACATGAGCTGCTACC 1320  
DB 1261 AGTGGCTCTCTGAACTCACTTCTCTGATCTGCAAAATGGGAAACATGAGCTGCTACC 1320  
QY 1321 TCTTAGACATGTGTGAGGAGACATATGATATTAACATGTATGTAAATCTCAGTGAT 1380  
DB 1321 TCTTAGACATGTGTGAGGAGACATATGATATTAACATGTATGTAAATCTCAGTGAT 1380  
QY 1381 GTCATGTAGGCTTAAACAGTGGGTGTGAGTCTGACTTAAAGTTACTGTTGCTG 1440  
DB 1381 GTCATGTAGGCTTAAACAGTGGGTGTGAGTCTGACTTAAAGTTACTGTTGCTG 1440  
QY 1441 AAAAAAAAAAAAAA 1454  
DB 1441 AAAAAAAAAAAAAA 1454

RESULT 6  
US-10-101-510-78  
; Sequence 78, Application US/10101510  
; Publication No. US20030148295A1  
; GENERAL INFORMATION:  
; APPLICANT: WAN, JACKSON  
; APPLICANT: WANG, YIXIN  
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE  
; FILE REFERENCE: 15117, 0012  
; CURRENT APPLICATION NUMBER: US/10/101, 510  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/276, 947  
; PRIOR FILING DATE: 2001-03-20  
; NUMBER OF SEQ. ID NOS: 805  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ. ID NO: 78  
; LENGTH: 1457  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-101-510-78

Query Match 99.2%; Score 1442; DB 12; Length 1457;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1453; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ACGCGCGCAGACCAAGCGAGGCGAGGCGAGCGTGGGCTCCCTGCTCTCTATC 60  
DB 1 ACGCGCGCAGACCAAGCGAGGCGAGGCGAGCGTGGGCTCCCTGCTCTCTATC 60  
QY 61 GCGGACTCCAGATCTGCGCATGAGAGCTCCGACCTCCACTCTCCGCGCTTGGC 120  
DB 61 GCGGACTCCAGATCTGCGCATGAGAGCTCCGACCTCCACTCTCCGCGCGCTTGGC 120  
QY 121 GCGCGGCTCTGGGAGAGCTGCTCCCTGCTGATGAGCGCAACTCTGGGCGGAGAGCG 180  
DB 121 GCGCGGCTCTGGGAGAGCTGCTCCCTGCTGATGAGCGCAACTCTGGGCGGAGAGCG 180  
QY 181 GCGCTGCTCCCGCAAAACGACACGCGCTTGAACCCCGAAGCTATGCGCGCGCG 240  
DB 181 GCGCTGCTCCCGCAAAACGACACGCGCTTGAACCCCGAAGCTATGCGCGCGCG 240  
QY 241 CCGGCGTCCAGAGCCCTGGAGAGCTGCTGCTTCAACGCGCTCTGCTTCACTGCGCGG 300  
DB 241 CCGGCGTCCAGAGCCCTGGAGAGCTGCTGCTTCAACGCGCTCTGCTTCACTGCGCGG 300  
QY 301 GTCCTGCTGAGACAGAGTGGTGGTCTGACGCGCGCGCACTGGGAAACAGCACTGT 360  
DB 301 GTCCTGCTGAGACAGAGTGGTGGTCTGACGCGCGCGCACTGGGAAACAGCACTGT 360  
QY 361 GCTCGAGTGGGGATGATCACTGCTGCTTCAAGGGGAGAGCTGCGCGGAGACT 420  
DB 361 GCTCGAGTGGGGATGATCACTGCTGCTTCAAGGGGAGAGCTGCGCGGAGACT 420



QY 742 GGCAGAGACCTTGGCCAGAGTGA CTGAGAGGCCCTGCTGTGACGAGACCTTCCA 801  
DB 661 GGCAGAGACCTTGGCCAGAGTGA CTGAGAGGCCCTGCTGTGACGAGACCTTCCA 720  
QY 802 GGCATCTCTGCTGGGGGTGTTTACCCCTGTGGCTCTGCGCCAGATCCAGCTGTCTACAC 861  
DB 721 GGCATCTCTGCTGGGGGTGTTTACCCCTGTGGCTCTGCGCCAGATCCAGCTGTCTACAC 780  
QY 862 CAGATCTGCAAAATACATGCTCTGATCAATAAAGTCAATAGCTCCAACTGA 912  
DB 781 CAGATCTGCAAAATACATGCTCTGATCAATAAAGTCAATAGCTCCAACTGA 831

## RESULT 8

US-09-964-824A-311  
; Sequence 311, Application US/09964824A  
; Patent No. US20020102531A1  
; GENERAL INFORMATION:  
; APPLICANT: Horrigan, Stephen  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signature  
; FILE REFERENCE: 689290-73  
; CURRENT APPLICATION NUMBER: US/09/964, 824A  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: US/60/236,033  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US/60/236,032  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US/60/236,028  
; PRIOR FILING DATE: 2000-09-28  
; NUMBER OF SEQ ID NOS: 583  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 311  
; LENGTH: 532  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-964-824A-311

Query Match 35.0%; Score 508.4; DB 10; Length 532;  
Best Local Similarity 99.4%; Pred. No. 1,8e-139;

Matches 531; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 712 AACACATGATATGTGCTGGA CTGAGACCGGGGCGGAGACCTTGGCCAGAGTGA 771  
DB 1 AACACATGATATGTGCTGGA CTGAGACCGGGGCGGAGACCTTGGCCAGAGTGA 60  
QY 772 GGGCCCTGCTGTGAGAGAGACCTTCCAAAGGCTCTGCTGGGGTGTTCACCTGT 831  
DB 61 GGGCCCTGCTGTGAGAGAGACCTTCCAAAGGCTCTGCTGGGGTGTTCACCTGT 120  
QY 832 GGGCTGCGCCAGATCCAGCTGTCTA CACCAGATCTGCAATACATGCTTGATCAAT 891  
DB 121 GGGCTGCGCCAGATCCAGCTGTCTA CACCAGATCTGCAATACATGCTTGATCAAT 179  
QY 892 AAAGTCATACGCTCCAACTGATCCAGATGCTACGCTCCAGCTGATCCAGATGTTATGCTC 951  
DB 180 AAAGTCATACGCTCCAACTGATCCAGATGCTACGCTCCAGCTGATCCAGATGTTATGCTC 239  
QY 952 CTGCTGATCCAGATGCGCCAGAGGCTCCATGCTCATCTCTTCTCTCCAGTGGCTGAA 1011  
DB 240 CTGCTGATCCAGATGCGCCAGAGGCTCCATGCTCATCTCTTCTCTCCAGTGGCTGAA 299  
QY 1012 CTCTCCCTTGTCTGCACTGTTCAAACCTTGCAGGCTCCAGACCTTAAACATCTCC 1071  
DB 300 CTCTCCCTTGTCTGCACTGTTCAAACCTTGCAGGCTCCAGACCTTAAACATCTCC 359  
QY 1072 CTCTCACTTCAATCCCACTATCCCACTATCTCTGCTCTGCTGCTGAAAGTGAATGCGAG 1131  
DB 360 CTCTCACTTCAATCCCACTATCCCACTATCTCTGCTCTGCTGCTGAAAGTGAATGCGAG 419  
QY 1132 AAGTGTGCGAAAGTGTATTCAGAGAGCGAGAGCGGCTATACCCAGCTCTGTA 1191

DB 420 AAGTGTGCGAAAGTGTATTCAGAGAGCGAGAGCGGCTATACCCAGCTCTGTA 479  
QY 1192 GAGCAGTTACTGCGGGGTCA CCAACTGACTTCTCTGCACTCCCGCTGTGTG 1245  
DB 480 GAGCAGTTACTGCGGGGTCA CCAACTGACTTCTCTGCACTCCCGCTGTGTG 532

## RESULT 9

US-09-867-701-5534  
; Sequence 5534, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Aglate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867, 701  
; PRIOR FILING DATE: 2001-05-29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5534  
; LENGTH: 532  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-867-701-5534

Query Match 35.0%; Score 508.4; DB 10; Length 532;  
Best Local Similarity 99.4%; Pred. No. 1,8e-139;

Matches 531; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 712 AACACATGATATGTGCTGGA CTGAGACCGGGGCGGAGACCTTGGCCAGAGTGA 771  
DB 1 AACACATGATATGTGCTGGA CTGAGACCGGGGCGGAGACCTTGGCCAGAGTGA 60  
QY 772 GGGCCCTGCTGTGAGAGAGACCTTCCAAAGGCTCTGCTGGGGTGTTCACCTGT 831  
DB 61 GGGCCCTGCTGTGAGAGAGACCTTCCAAAGGCTCTGCTGGGGTGTTCACCTGT 120  
QY 832 GGGCTGCGCCAGATCCAGCTGTCTA CACCAGATCTGCAATACATGCTTGATCAAT 891  
DB 121 GGGCTGCGCCAGATCCAGCTGTCTA CACCAGATCTGCAATACATGCTTGATCAAT 179  
QY 892 AAAGTCATACGCTCCAACTGATCCAGATGCTACGCTCCAGCTGATCCAGATGTTATGCTC 951  
DB 180 AAAGTCATACGCTCCAACTGATCCAGATGCTACGCTCCAGCTGATCCAGATGTTATGCTC 239  
QY 952 CTGCTGATCCAGATGCGCCAGAGGCTCCATGCTCATCTCTTCTCTCCAGTGGCTGAA 1011  
DB 240 CTGCTGATCCAGATGCGCCAGAGGCTCCATGCTCATCTCTTCTCTCCAGTGGCTGAA 299  
QY 1012 CTCTCCCTTGTCTGCACTGTTCAAACCTTGCAGGCTCCAGACCTTAAACATCTCC 1071  
DB 300 CTCTCCCTTGTCTGCACTGTTCAAACCTTGCAGGCTCCAGACCTTAAACATCTCC 359  
QY 1072 CTCTCACTTCAATCCCACTATCCCACTATCTCTGCTCTGCTGCTGAAAGTGAATGCGAG 1131  
DB 360 CTCTCACTTCAATCCCACTATCCCACTATCTCTGCTCTGCTGCTGAAAGTGAATGCGAG 419  
QY 1132 AAGTGTGCGAAAGTGTATTCAGAGAGCGAGAGCGGCTATACCCAGCTCTGTA 1191  
DB 480 AAGTGTGCGAAAGTGTATTCAGAGAGCGAGAGCGGCTATACCCAGCTCTGTA 1245

## RESULT 10

US-09-954-531-611  
; Sequence 611, Application US/09954531  
; Patent No. US20020165180A1





Db 33 ACTATGATATTAACATGTGTATGTAATCTTCA 2  
Search completed: November 25, 2003, 09:06:16  
Job time : 491 secs

QY 1214 ACCCTGACTTCTCTGCGCACTCCCGCTGTGTACTTTGGGCAAGCAAGTCCCTCTCTG 1273  
DB 230 ACCCTGACTTCTCTGCGCACTCCCGCTGTGTACTTT-GGCAAGCAAGTCCCTCTCTG 172  
QY 1274 AACCTGATTTCTCTCTGCGCACTCCCGCTGTGTACTTTGGGCAAGCAAGTCCCTCTCTG 1333  
DB 171 AACCTGATTTCTCTCTGCGCACTCCCGCTGTGTACTTTGGGCAAGCAAGTCCCTCTCTG 112  
QY 1334 GTGAGGAGACTATGATATTAACATGTGTATGTAATCTTCAAT-GTATGTCATGTAAAGC 1392  
DB 111 GTGAGGAGACTATGATATTAACATGTGTATGTAATCTTCAATGTATGTATGTATGTAAGC 52  
QY 1393 TTTACACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1443  
DB 51 TTTACACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1

## RESULT 15

US-09-867-701-2781/C  
Sequence 2781, Application US/09867701  
Patent No. US20020132237A1  
GENERAL INFORMATION:  
APPLICANT: Agiate, Paul A.  
APPLICANT: Jones, Robert  
APPLICANT: Harlocker, Susan L.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.497  
CURRENT APPLICATION NUMBER: US/09/867,701  
NUMBER OF SEQ ID NOS: 10912  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2781  
LENGTH: 468  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-867-701-2781

Query Match 30.2%; Score 438.4; DB 10; Length 468;  
Best Local Similarity 99.6%; Pred. No. 7,7e-119;  
Matches 450; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 922 TACGCTCCAGCTGATCCAGATGTTATGCTCTGCTGATCCAGATGCCAGAGCTCCATC 981  
DB 452 TACGCTCCAGCTGATCCAGATGTTATGCTCTGCTGATCCAGATGCCAGAGCTCCATC 393  
QY 982 GTCCATCT 1041  
DB 392 GTCCATCT 333  
QY 1042 TGCAGCTCTCCAGCTGATCCAGATGTTATGCTCTGCTGATCCAGATGCCAGAGCTCCAT 1101  
DB 332 TGCAGCTCTCCAGCTGATCCAGATGTTATGCTCTGCTGATCCAGATGCCAGAGCTCCAT 273  
QY 1102 TCTTCTGCTGATCTGAGCTGAAGTCAAGTGGTGGCAAGGTTTATCCAGAGAG 1161  
DB 272 TCTTCTGCTGATCTGAGCTGAAGTCAAGTGGTGGCAAGGTTTATCCAGAGAG 213  
QY 1162 CCAGAGAGCTGATCCAGCTCTGAGAGAGTTACTGGGGTCAACCAACTGACT 1221  
DB 212 CCAGAGAGCTGATCCAGCTCTGAGAGAGTTACTGGGGTCAACCAACTGACT 153  
QY 1222 TCTTCTGCTGATCTGAGCTGAAGTGGGCAAGGTTTATCCAGAGAG 1281  
DB 152 TCTTCTGCTGATCTGAGCTGAAGTGGGCAAGGTTTATCCAGAGAG 94  
QY 1282 TTTCTCTCATCTGCAAAATGGGAAACATGACGTGCTTACATGTTATGAGAGAG 1341  
DB 93 TTTCTCTCATCTGCAAAATGGGAAACATGACGTGCTTACATGTTATGAGAGAG 34  
QY 1342 ACTATGATATTAACATGTGTATGTAATCTTCA 1373





QY	6	GGGGAATCCCAAGATCCCTGGGCAATGAGAGCTCGGCACTCCACCTCTCCGAGCGCTCTGGC	120
Db	61	GGGGAATCCCAAGATCCCTGGGCAATGAGAGCTCGGCACTCCACCTCTCCGAGCGCTCTGGC	120
QY	121	GGCCGGGACTCTGGCGAAGCTGCTGCGGCTGTGATGATGCGCAACTCTGGGCGCGAAGAGCGG	180
Db	121	GGCCGGGACTCTGGCGAAGCTGCTGCGGCTGTGATGATGCGCAACTCTGGGCGCGAAGAGCGG	180
QY	181	GGGCTGCTCTCCCGAAAAAGACACAGCGGCTTGGACCCCGAAGCCTATGCGCGCCCTGTGCGG	240
Db	181	GGGCTGCTCTCCCGAAAAAGACACAGCGGCTTGGACCCCGAAGCCTATGCGCGCCCTGTGCGG	240
QY	241	CGCGGCTCGAGAGCCCTGGGAGGCTGTGGCTCTTCAACAGGCTCGGTTCGACTGCGGAGGCT	300
Db	241	CGCGGCTCGAGAGCCCTGGGAGGCTGTGGCTCTTCAACAGGCTCGGTTCGACTGCGGAGGCT	300
QY	301	GTCTGTGTGGACCAAGATTGGGTGTGACGCGCTCGGCACTGCGGAAAAACAGGCATGTGG	360
Db	301	GTCTGTGTGGACCAAGATTGGGTGTGACGCGCTCGGCACTGCGGAAAAACAGGCATGTGG	360
QY	361	GCTCGAGTAGGGAGATGATCACTGGCTGTCTTCAAGGGGAGACAGTCCGCGGAGCAT	420
Db	361	GCTCGAGTAGGGAGATGATCACTGGCTGTCTTCAAGGGGAGACAGTCCGCGGAGCAT	420
QY	421	CGCTCTGTGTGTCATATCCCAAGTACCACAGAGGCTCAAGGCGCATCTGTGCAAGAGCGAAG	480
Db	421	CGCTCTGTGTGTCATATCCCAAGTACCACAGAGGCTCAAGGCGCATCTGTGCAAGAGCGAAG	480
QY	481	GATGAGGACGATCTCATGTTGTCTAAAGCTGGGCAAGGCGCGTATGTCGCGGAGCTCCGCGCT	540
Db	481	GATGAGGACGATCTCATGTTGTCTAAAGCTGGGCAAGGCGCGTATGTCGCGGAGCTCCGCGCT	540
QY	541	CGGAGCCCTGACACTTCCCTACCGCTGTGTGTACGCGCGAGACAGACAGTGCAGAGTGTGGC	600
Db	541	CGGAGCCCTGACACTTCCCTACCGCTGTGTGTACGCGCGAGACAGACAGTGCAGAGTGTGGC	600
QY	601	TGGGGCAACCAAGGCGCGCGCGGAGTGAAGTCAACAGAGGCTTGACTCTCTCCAGATC	660
Db	601	TGGGGCAACCAAGGCGCGCGCGGAGTGAAGTCAACAGAGGCTTGACTCTCTCCAGATC	660
QY	661	ACTATCTGAGGCTCTAAAGAGTGTAGAGTCTTCAACCTCGGCGTGTATCCAAACAACATG	720
Db	661	ACTATCTGAGGCTCTAAAGAGTGTAGAGTCTTCAACCTCGGCGTGTATCCAAACAACATG	720
QY	721	ATATGTGTGGAATCTGAGACCTGGGAGCAAGACCTTTCAGAGTGAATCTTGAAGAGGCGCCCTG	780
Db	721	ATATGTGTGGAATCTGAGACCTGGGAGCAAGACCTTTCAGAGTGAATCTTGAAGAGGCGCCCTG	780
QY	781	GTCTGTGAGAGAACCTTCGAAAGGACATCTGTGTGGGGTGTAAACCTGTGGCTGTGCG	840
Db	781	GTCTGTGAGAGAACCTTCGAAAGGACATCTGTGTGGGGTGTAAACCTGTGGCTGTGCG	840
QY	841	CAGCATCCAGCTGTGTACACCCAGATCTGCAATACATGTCTCTGATCAATAAATCATTA	900
Db	841	CAGCATCCAGCTGTGTACACCCAGATCTGCAATACATGTCTCTGATCAATAAATCATTA	900
QY	901	CGCTCCCACTGATCCAGATGCTAGAGTCACTGAGTGTATCCAGATGTTATGCTCTGTGTATC	960
Db	901	CGCTCCCACTGATCCAGATGCTAGAGTCACTGAGTGTATCCAGATGTTATGCTCTGTGTATC	960
QY	961	CAGATGCCAGAGGCTCATGCTCATCTCTTCCCTCCAGTGGGCTGAACTCTCCCT	1020
Db	961	CAGATGCCAGAGGCTCATGCTCATCTCTTCCCTCCAGTGGGCTGAACTCTCCCT	1020
QY	1021	TGTCTGCACTGTTCAACCTCTGCGCGCTTCACACTCTTAAACATCTCCCTCTCACCT	1080
Db	1021	TGTCTGCACTGTTCAACCTCTGCGCGCTTCACACTCTTAAACATCTCCCTCTCACCT	1080
QY	1081	CATTCCCCCACTATCCCCATTTCTGTGCTGTATCTGAACTGAAATGACGAAAGGAGTGG	1140
Db	1081	CATTCCCCCACTATCCCCATTTCTGTGCTGTATCTGAACTGAAATGACGAAAGGAGTGG	1140
QY	1141	CAAAAGTTTATTCAGAGAGCCAGAAAGCCGAGTATCAACCGACTCTGAGAGCATTA	1200

Db	1141	CAAGGTTATTCCAGAGAAAGCCAGGAAGCCGGCTGATCATCCAGACCTCTGAGGACAGTTA	1200
QY	1201	CTGGGGGTACCCCAACCTGACTTCTCTCTGACCCTCCCCGTGTGTACTTTGGGCAAGCCA	1260
Db	1201	CTGGGGGTACCCCAACCTGACTTCTCTCTGACCCTCCCCGTGTGTACTTTGGGCAAGCCA	1260
QY	1261	AGTGGCCCTCTGTGAACCTCAGTTTCCCTCATCTGCAAAATGGGAAACAATGAGAGTCCATCC	1320
Db	1261	AGTGGCCCTCTGTGAACCTCAGTTTCCCTCATCTGCAAAATGGGAAACAATGAGAGTCCATCC	1320
QY	1321	TCTTGAACATGTTGTGAGAGACATGATATTAACATGTGTATGTAAATCTTCATGTGATT	1380
Db	1321	TCTTGAACATGTTGTGAGAGACATGATATTAACATGTGTATGTAAATCTTCATGTGATT	1380
QY	1381	GTCATGTGAAGCCTTAACAACAGTGGGTGGTGAAGTTCTGACTAAAGTTAACTGTTGTGCTG	1440
Db	1381	GTCATGTGAAGCCTTAACAACAGTGGGTGGTGAAGTTCTGACTAAAGTTAACTGTTGTGCTG	1440
QY	1441	AAAAAAAAAAAAAAAA 1454	
Db	1441	AAAAAAAAAAAAAAAA 1454	

RESULT 2  
 US-08-628-198-2  
 ; Sequence 2, Application us/08628198  
 ; Patent No. 5843594  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Band, Yvmla  
 ; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
 ; TITLE OF INVENTION: MOLECULES AND METHODS  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 225 Franklin Street  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02110-2804  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: us/08/628,198  
 ; FILING DATE:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/467,155  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Clark, Paul T.  
 ; REGISTRATION NUMBER: 30,162  
 ; REFERENCE/DOCKET NUMBER: 00398/100002  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617/542-8070  
 ; TELEFAX: 617/542-8906  
 ; TELEX: 200154  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1454 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; US-08-628-198-2

Query Match	100.0%	Score 1454;	DB 2;	Length 1454;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1454; Conservative	0;	Mismatches	0;	Gaps 0;









TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
 TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
 TITLE OF INVENTION: OF THE PROSTATE  
 NUMBER OF SEQUENCES: 76  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Abbott Laboratories  
 STREET: 100 Abbott Park Road  
 CITY: Abbott Park  
 STATE: IL  
 COUNTRY: USA  
 ZIP: 60064-3500  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/944,483  
 FILING DATE:  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Becker, Cheryl L.  
 REGISTRATION NUMBER: 35,441  
 REFERENCE/DOCKET NUMBER: 6183.US.01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 847/935-1729  
 TELEFAX: 847/938-2623  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1192 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-08-944-483-8

Query Match	Similarity	11.3%	Score 164	DB 3	Length 1192
Best Local	Similarity	55.1%	Pred No.5,2e-33		
Matches	370	Conservative	0	Mismatches 290	Indels 12
					Gaps 2
QY	222	CTATGCGCGCCCGCTGCGCGCGCGGGCTGCGAGCCCTGCGAGAGTCTGCTCTTCAACGAGCCT	281		
Db	175	CAAGGGGTTCGAGTCCAAAGCTCACTCCAGCGCCTTGCGAGCGAGCGCCCTTTGAGAAAGAC	234		
QY	282	CTCGTTCCACAGCGCGGGGCTGCTCTGCTGAGACCGAGGTTGGGTGCTGACGCGCCGCGACTG	341		
Db	235	GCGGTCACTCTGTGCGGGCGAGCGCTATAGCCCCCGAGTGGCTCTCGAAGCGAGCCACTG	294		
QY	342	CGAAACAAGCCACTGTGGGCTCGAGTAGAGGGGATATACCTGTCTTCTCAGGGCGA	401		
Db	295	CCTCAAGCCCGCTACATGATTCACCTCGGGGAGCAACAACCTC-----CAAGAAGA	345		
QY	402	GCAGTCCCGCGGAGCGACTCGCTGTGTGTCATCCCAATGACCAAGGCTCAAGGCC	461		
Db	346	GGAAGGCTGTAGACAGACCCGAGCGGCACCTGAAGTCTTCCCGCCACCCCGGTTTCAACA	405		
QY	462	CATCTCGSCAAGCGGAGAGCGATGAGACAGATCTCATGTTGCTTAAAGTGGCGAGGCCGCT	521		
Db	406	CAGCTCTCCCAACAAGACCAACCGAATGACATCATGTGTGATGAAGATGGCATGGCCAGT	465		
QY	522	AGTCCCGGGGCCCCCGCTCCGGGCGCTGCAAGTTCCCTAACCGTGTGCTCAAGCCCGAGAGA	581		
Db	466	CTTCATCACTCGGAGCTGTGCGAACCCCTCAACCTCTTCCACAGCGTGTGTCACGTGTGGAC	525		
QY	582	CCAAGTCCAGAGTTGCTGGCTGGGCGACCAAGCGCGCCCGAGAGAGGAATCAACAAGGG	641		
Db	526	CAGTGGCTCATTTTCCGGCTGGGGCAGACGCTCAGGCCCCCAAGTTAGCCCTGCTCAAC	585		
QY	642	CTGACCTGCTCCAGCATCACTATCTTGAAGCCCTTAAAGGTGAGAGTCTTCTAACCTCGG	701		

Db	586	CTTGGAGATGGCGCAATCATCACCATATGAGCAACCAAGATGTGAGAAACGCTTAACCCCGG	645
QY	702	CGTGTCAACCAACCATGATATGTC---TGGACTGGACCGGGCCAGAACCTTTGGCA	758
Db	646	CACATCAACAGACACCACTAGTGTGTGTGCCACGCTGAGAGAGGGGGGAGAGACTCTGCGCA	705
QY	759	GAGTGACTCTGAGAGGCCCCCTGTGTCTGTGACGAGAACCTTCACAGGCATTCCTCTGTGGG	818
Db	706	GAGTGAATCCCGGGGGCCCTCTGTGTCTGTGAACCAAGCTCTTCACAGGATATATCTCCGTGGG	765
QY	819	TGTTTAAACCTGTGGCTCTGGCCAGCATCCAGCTGTCTACACCCAGATTTGCAAAATACAT	878
Db	766	CCAGGATTCGGTGTGGATCAACCCCAAAACCTGTGTCTACACGAAAGTTCGCAAAATATGT	825
QY	879	GTCCCTGATCAA	890
Db	826	GGACTGGATCCA	837

```

1      RESULT 8
2      US-09-025-059-2
3      : Sequence 2, Application US/09025059
4      : Patent No. 6075136
5      : GENERAL INFORMATION:
6      : APPLICANT: Tang, Y. Tom
7      : APPLICANT: Corley, Neil C.
8      : APPLICANT: Guesler, Karl J.
9      : TITLE OF INVENTION: PROSTATE-ASSOCIATED SERINE PROTEASE
10     : NUMBER OF SEQUENCES: 4
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: Incyte Pharmaceuticals, Inc.
13     : STREET: 3174 Porter Dr.
14     : CITY: Palo Alto
15     : STATE: CA
16     : COUNTRY: USA
17     : ZIP: 94304
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Diskette
20     : COMPUTER: IBM Compatible
21     : OPERATING SYSTEM: DOS
22     : SOFTWARE: FastSeq for Windows Version 2.0
23     : CURRENT APPLICATION DATA:
24     : APPLICATION NUMBER: US/09/025,059
25     : FILING DATE:
26     : CLASSIFICATION:
27     : PRIOR APPLICATION DATA:
28     : APPLICATION NUMBER:
29     : FILING DATE:
30     : ATTORNEY/AGENT INFORMATION:
31     : NAME: Billings, Lucy J
32     : REGISTRATION NUMBER: 36,749
33     : REFERENCE/DOCKET NUMBER: PF-0481 US
34     : TELECOMMUNICATION INFORMATION:
35     : TELEPHONE: 650-855-0555
36     : TELEFAX: 650-845-4166
37     : TELEX:
38     : INFORMATION FOR SEQ ID NO: 2:
39     : SEQUENCE CHARACTERISTICS:
40     : LENGTH: 1314 base pairs
41     : TYPE: nucleic acid
42     : STRANDEDNESS: single
43     : TOPOLOGY: linear
44     : IMMEDIATE SOURCE:
45     : LIBRARY: LUNGJUT010
46     : CLONE: 2723646
47     : US-09-025-059-2

```

Query Match	11.3%	Score 164	DB 3	Length 1314
Best Local Similarity	55.1%	Pred. No. 5.5e-33		
Matches 370	Conservative 0	Mismatches 290	Indels 12	Gaps 2

222 CTAATGAGGCCCGATGGCGGCGGGGATGCGAGCCCTGGCAGGCTTCGCTTTCAACGGCCT 281

Db 292 CAAGGGGCTTGAGTGCAGAGCTCACTCCAGCCCTGGCAGGAGCCCTGTTCAGAAAGC 351  
Qy 282 CTGTTTCACTGCGCGGGGTGCTCGTGGAGCAGAGTGGGTCTTAACGGCCCGGACATG 341  
Db 352 GCGGCTACTGTGGGGGCGACGCTCATCGCCCCAGATGGCTCTTGAACAGACCCACATG 411  
Qy 342 CCGAAMCAAGCCACTGTGGGCTCGAGTGGGGATGATCACTGCTCTTCTTCAAGGGCGA 401  
Db 412 CTTCAAGCCCCGCTACATAGTTACCTGGGGGAGCAACACCTC-----CAGAAAGH 462  
Qy 402 GCGAGCTCCCGGAGACACTGCTCTGTTCATCCCAAGTACCAAGGCTCAGGGCCC 461  
Db 463 GAGGGGCTGGAGCAGAGCCCGGACAGCAGCTGAGTCTTCCACCCGGCTTCAACAA 522  
Qy 462 CATCTGCCAAGCGCAAGGATAGACAGATCATGTTGCTTAAGCTGGCCGAGCCGCT 521  
Db 523 CACCTCTCCCAAAAGAACCCGCGCAATGATCATGCTGGTGAAGATGGCATCGCCAGT 582  
Qy 522 AGTCCGGGGCCCCCGGCTCGGGGCTCGAGCTTCCCTAACCGCTGTCTCAGCCCGAGA 581  
Db 583 CTCATCACTGAGCTGTGCGACCTCCACCTCTCTCTCAAGCTGTGTCATGCTGGAC 642  
Qy 582 CCAAGTGCAGGTTGCTGGCTGGGGGCAACAGCGCCCGGAGAGTGAATACAAAGG 641  
Db 643 CAGCTGCTCATTTCCGCTGGGGGAGCAGAGCTCAGCCCCAGTTACGCTGCTCAC 702  
Qy 642 CTGACCTGCTCCAGATCATATCTGAGCCCTTAAAGATGTGAGGTCTTACCTG 701  
Db 703 CTTCGATGGCGCAATCATCAATGATGAGCAGCAAGAGTGTGAAGAGCCCTACCCGG 762  
Qy 702 CCGTGTACCAAAATATATATGTC---TGACGTGACCGGGGGCAAGAGCCCTTGCA 758  
Db 763 CAACATCAAGACACATGATGTGTCCAGCGTGCAGAGGGGGCAAGAGATCTCCGCA 822  
Qy 759 GAGTGACTGAGAGCCCTGCTGTGTGACGAGACCTCCCAAGGATCTCTCGGGG 818  
Db 823 GGGTGACTCCGGGGGCTCTGTCTGTAAACAGATCTCTTCAAGGATATATCTCTGGG 882  
Qy 819 TGTTTTCCCTGTGTCTGCTGCGGAGATCATGCTTACCCAGATCTGCAAAATACAT 878  
Db 883 CCAAGATCCGTGTGCTGATCACCAGGAGCTGTGTCTACAGAAAGTCTGCAAAATATG 942  
Qy 879 GTCCTGATCAA 890  
Db 943 GGACTGATCCA 954

RESULT 9  
US-09-205-189  
; Sequence 189, Application US/09205258  
; Patent No 6525174  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 207 Human Secreted Proteins  
; FILE REFERENCE: P2007P1  
; CURRENT APPLICATION NUMBER: US/09/205,258  
; EARLIER FILING DATE: 1998-12-04  
; EARLIER APPLICATION NUMBER: PCT/US98/11422  
; EARLIER FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,885  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,375  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,881  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,880  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,896  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,020  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,876

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,895  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,884  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,894  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,971  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,964  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,882  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,899  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,893  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,900  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,901  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,892  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,915  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,019  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,970  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,972  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,916  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,373  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,875  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,374  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,917  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,949  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,974  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,883  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,897  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,898  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,962  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,963  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,877  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,878  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/070,923  
; EARLIER FILING DATE: 1997-12-18  
; EARLIER APPLICATION NUMBER: 60/092,921  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/094,657  
; EARLIER FILING DATE: 1998-07-30  
; NUMBER OF SEQ ID NOS: 1227  
; SOFTWARE: Patent Ver. 2.0  
; SEQ ID NO 189  
; LENGTH: 1292  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-205-258-189

QY	222	CTATGACGCGCCCGTGGCCCGCGGCGGTCTGCGAGCGCCCTGCGAGAGTCTTGGCTCTTCAACGGCGCT	281
Db	275	CAAGGGAGTTGCGAGTGCMAAGCCCTCACTCCAGCGCTGCGAGGACAGCCCTGTTCCGAGAAAGC	334
QY	282	CTGTGTTGCACCTGCGCGGAGTGTCTCGGTGGAGTCCAGAGTGGAGTGGTGGAGCGCGCGGCACTG	341
Db	335	GCGGCTACTCTGTGGGGGCGACGCTATTCGCCCGCCAGATGGCTCTTGAAGAGAGCCCACTG	394
QY	342	CGGAAACAAAGCCACTGTGGGCTCTCGAGTAGGGAGATCACTGCTGCTTCTTCAGGGCGA	401
Db	395	CTTCMAAGCCCCCGTACATAGTTTCACCTGGGGGAGCAACAACCTC-----CAGAAAGA	445
QY	402	GCAAGTCTCGCGCGGACACTGCTGTGTGTCTTCATCCCAAGTACACCAAGGCGCTCAGGCC	461
Db	446	GAGAGGCTGTGAGCAGACCCCGAGACGCCATGAGTCTTCCGCCACCCCGGCTTCAGAA	505
QY	462	CATCTTGGCCAAAGCGCAAGGATGAGCAGATCTCATGTTCCTAAAGCTTGGCTGCGCAAGCCCT	521
Db	506	CACCTCTCCCAACAAAGACACCGCAATACATACCTGTGTGAAGAGTGGATGCGCAGT	565
QY	522	AGTGCAGGGGCCCCCGGCTCCGGGCCCCGTGAGCTTCCCTACACGCTGTGTGCTAGCCCCGAGA	581
Db	566	CTCATATCACCTGGAGTGTGCGAACCCCTCACCTCTCTCCAGCTGTGTCACTGTGTGGAC	625
QY	582	CCAGTGCACAGTTGTGCTGTGGCTGGAGCACACCGCGCGCGCCGAGAGAGTAGTACAAACAAGG	641
Db	626	CAGCTGTCTCATTTTCCGGCTGGGGGACAGCTGTCCAGCCCCCAGTTACGCCCTGCTCAAC	685
QY	642	CTGACCTGCTCCAGCATCACTATCTTGAACCTTAAGAGTGTGAAGTCTTCTACCCCTGG	701
Db	686	CTTGCATATGGCCCAATCATCATATGAGACACGAGAGTGTGAACCGCTTACCCCGG	745
QY	702	CGTGTACCAACAATCATGATATGTGC--TGAATGAGACCGGGGCGCAGACCCCTTSCCA	758
Db	746	CACATCTACAGACACATGTGTGTGTGCCGCTGAGGAAAGGGGCAAGACTCTCTGCCA	805
QY	759	GAGTGACTCGAGAGCGCCCTGTGTCTGTGAAGAGACCTCCAAAGGCATCTCTCTGTGGG	818
Db	806	GGGTGTACTCGGGGGGCCCCCTGTGTGTGTGAACAGTCTTTCAAGGCATTAATCTCTGGGG	865
QY	819	TGTTTACCCCTGTGGCTGTGCGCAGATCCAGCTTCTACACCCAGATCTGCAATACAT	878
Db	866	CGAGATTCGTGTGTGGCATCACCGGAAAGCTGTGTCTTACAGAAAGTCTGCAATATGT	925
QY	879	GTCTGTGATCAA	890
Db	926	GGACTGGATCA	937

US-RESULT 10  
 US-09-205-258-247  
 Sequence 247, Application US/09205258  
 Patent No. 6525174  
 GENERAL INFORMATION:  
 APPLICANT: Young et al.  
 TITLE OF INVENTION: 207 Human Secreted Proteins  
 FILE REFERENCE: P2007071  
 CURRENT APPLICATION NUMBER: US/09/205.258  
 CURRENT FILING DATE: 1998-12-04  
 EARLIER APPLICATION NUMBER: PCT/US98/11422  
 EARLIER FILING DATE: 1998-06-04  
 EARLIER APPLICATION NUMBER: 60/048.885  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/049.375  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048.881  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048.880  
 EARLIER FILING DATE: 1997-06-06

1	EARLIER APPLICATION NUMBER: 60/048,896
2	EARLIER FILING DATE: 1997-06-06
3	EARLIER APPLICATION NUMBER: 60/049,020
4	EARLIER FILING DATE: 1997-06-06
5	EARLIER APPLICATION NUMBER: 60/048,876
6	EARLIER FILING DATE: 1997-06-06
7	EARLIER APPLICATION NUMBER: 60/048,895
8	EARLIER FILING DATE: 1997-06-06
9	EARLIER APPLICATION NUMBER: 60/048,888
10	EARLIER FILING DATE: 1997-06-06
11	EARLIER APPLICATION NUMBER: 60/048,892
12	EARLIER FILING DATE: 1997-06-06
13	EARLIER APPLICATION NUMBER: 60/048,899
14	EARLIER FILING DATE: 1997-06-06
15	EARLIER APPLICATION NUMBER: 60/048,899
16	EARLIER FILING DATE: 1997-06-06
17	EARLIER APPLICATION NUMBER: 60/049,019
18	EARLIER FILING DATE: 1997-06-06
19	EARLIER APPLICATION NUMBER: 60/048,977
20	EARLIER FILING DATE: 1997-06-06
21	EARLIER APPLICATION NUMBER: 60/048,972
22	EARLIER FILING DATE: 1997-06-06
23	EARLIER APPLICATION NUMBER: 60/048,916
24	EARLIER FILING DATE: 1997-06-06
25	EARLIER APPLICATION NUMBER: 60/049,373
26	EARLIER FILING DATE: 1997-06-06
27	EARLIER APPLICATION NUMBER: 60/048,878
28	EARLIER FILING DATE: 1997-06-06
29	EARLIER APPLICATION NUMBER: 60/049,374
30	EARLIER FILING DATE: 1997-06-06
31	EARLIER APPLICATION NUMBER: 60/048,911
32	EARLIER FILING DATE: 1997-06-06
33	EARLIER APPLICATION NUMBER: 60/048,945
34	EARLIER FILING DATE: 1997-06-06
35	EARLIER APPLICATION NUMBER: 60/048,974
36	EARLIER FILING DATE: 1997-06-06
37	EARLIER APPLICATION NUMBER: 60/048,883
38	EARLIER FILING DATE: 1997-06-06
39	EARLIER APPLICATION NUMBER: 60/048,897
40	EARLIER FILING DATE: 1997-06-06
41	EARLIER APPLICATION NUMBER: 60/048,898
42	EARLIER FILING DATE: 1997-06-06
43	EARLIER APPLICATION NUMBER: 60/048,898
44	EARLIER FILING DATE: 1997-06-06
45	EARLIER APPLICATION NUMBER: 60/048,878
46	EARLIER FILING DATE: 1997-06-06
47	EARLIER APPLICATION NUMBER: 60/070,922
48	EARLIER FILING DATE: 1997-12-18
49	EARLIER APPLICATION NUMBER: 60/092,922
50	EARLIER FILING DATE: 1998-07-15
51	EARLIER APPLICATION NUMBER: 60/094,655
52	EARLIER FILING DATE: 1998-07-30
53	NUMBER OF SEQ ID NOS: 1227
54	SOFTWARE: PatentIn Ver. 2.0
55	SEQ ID NO 247



```

; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (20)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (35)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (36)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (37)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-205-258-247
```

Query Match 11.2%; Score 163.2; DB 4; Length 1146;  
Best Local Similarity 54.8%; Pred. No. 8,3e-33;  
Matches 368; Conservative 2; Mismatches 290; Indels 12; Gaps 2;

```

QY 222 CTATGCGCCCGGCTGCGCGCGCGCTGCGAGCCCTGCGAGTCTGCTTCAACGCGCT 281
DB 129 CAAGGGGTTGAGTGAAGCTTCACTCCAGCCCTGCGAGCGCGAGCCCTGCTGAGAGAC 188
QY 282 CTGCTTCACTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 341
DB 189 GGGGCTACTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 248
QY 342 CGGAAACAAGCACTGTGCGCGCTGCGAGTGAATGATCACTGCTTCTTCAAGCGCGA 401
DB 249 CTTCAAGCCCGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 299
QY 402 GCGAGTCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 461
DB 300 GAGAGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359
QY 462 CATCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 521
DB 360 GAGCGTCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
QY 522 AGTGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 581
DB 420 CTCATCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 479
QY 582 CCAGTGCAGGCTGTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 641
DB 480 GAGCTGTCTCATTTCCGCGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 539
QY 642 CTTGACCTGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 701
DB 540 CTTGGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
QY 702 CGTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 758
DB 600 CAACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659
QY 759 GAGTACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 818
DB 660 GAGTACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
QY 819 TGTTCACCTGTGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 878
DB 720 CGAGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
QY 879 GTCTCTGATCA 890
DB 780 GAGCTGATCA 791
```

```

RESULT 11
US-09-386-642-10
; Sequence 10, Application US/09386642
; Patent No. 6420157
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jenson
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Zymogen Activation System
; FILE REFERENCE: ORT-1028
; CURRENT APPLICATION NUMBER: US/09/386,642
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1052
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
; OTHER INFORMATION: With homo sapien serine protease catalytic domain
US-09-386-642-10
```

Query Match 11.1%; Score 160.8; DB 4; Length 1052;  
Best Local Similarity 55.4%; Pred. No. 3,4e-32;  
Matches 359; Conservative 0; Mismatches 277; Indels 12; Gaps 2;

```

QY 246 CTCGAGCGCTGCGAGGCTGCTGCTTCAACGCGCTGCTGCTGCTGCTGCTGCTGCT 305
DB 198 CTCGAGCGCTGCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 257
QY 306 GATGAGCCAGAGTGGGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 365
DB 258 CATGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 317
QY 366 AGTGGGAGATGATCACTGCTGCTTCTCAAGGCGAGAGAGAGAGAGAGAGAGAGAGAG 425
DB 318 CTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 368
QY 426 TGTGTCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 485
DB 369 AGCACTGAGTCTTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428
QY 486 GCAAGATCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 545
DB 429 CAATGACATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 488
QY 546 CTTGAGCTTCTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 605
DB 489 CTTCACTCTCTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 548
QY 606 CACAGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 665
DB 549 CAGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 608
QY 666 CTTGAGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 725
DB 609 CATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 668
QY 726 TGC---TGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 782
DB 669 TGCAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 728
QY 783 CTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 842
DB 729 CTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 788
QY 843 GCATCAAGCTGTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 890
DB 789 AAGCGTGTGTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 836
```

## RESULT 12

US-08-790-137-2  
 ; Sequence 2, Application US/08790137  
 ; Patent No. 5840871  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hillman, Jennifer L.  
 ; TITLE OF INVENTION: A NOVEL PROSTATE-ASSOCIATED  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; OPERATING SYSTEM: DOS Compatible  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/790,137  
 ; FILING DATE: Filed Herewith  
 ; CLASSIFICATION: 424  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0195 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-855-0555  
 ; TELEFAX: 415-845-4166  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 833 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-790-137-2

Query Match 11.0%; Score 160.2; DB 2; Length 833;  
 Best Local Similarity 54.8%; Pred. No. 4,4e-32;  
 Matches 362; Conservative 0; Mismatches 286; Indels 12; Gaps 2;

QY 234 GTGCGGCGCGGCTGCGACGCTGCGAGTCTGCTTCAACGCGCTCTGCTCACTG 293  
 DB 116 GTGTGAGCAGATCTCCAGCGCTGAGCGGCTCTGTACGAAAGACGCGCTACTG 175  
 QY 294 CGCGGCTGCTGTGTGACCAAGTTGAGTGTGACGCGCGGCGCACTGCGAAACAAAGCC 353  
 DB 176 TGGGGGAGAGNATCATGAGCCCGGAGATGTTCTGTAGAGAGCCACTCTCCTMAAGCCCG 235  
 QY 354 ACTGTGGCTCTGAGTATGAGGAGATGATCACTGCTGCTTCTTCAAGGCGAGAGCTCCGCG 413  
 DB 236 CTACATATGTTCACTGGGCGACACACACTC-----CAGAGGAGAGAGGCTGTGA 286  
 QY 414 GACGATCTGCTGTGTGCTCATCCCAAGTACACAGGAGCTGAGCGCCCATCTGCGCAAG 473  
 DB 287 GCGAGCGCGAGACAGACAGATGATCTCTCCCGGCGCTTCAACAAAGAGCTCCCA 346  
 QY 474 GCGAAGCAGTACAGCAGATCTATGTTGTAAGCTGCGCAGCGCCGTAAGTGGCGGAGCC 533  
 DB 347 CAAGAGCAGCAGCAGATGATCATCTGTGTGAAAGTGGATGCGCAGTCTCATACACTG 406  
 QY 534 CGCGGCGCGGCGCTGAGCTTCCCTACCGCTGTGCTGAGCGCGGAGCAGAGTGGCAAGT 593  
 DB 407 GGGTGTGCGACCCCTCACTCTCTCTCAAGCTGTGTGCTGAGCAGCAGTGTCTAT 466  
 QY 594 TGTGCTGTGGGACACCGCGCGCGGAGAGTGAAGTAAACAAAGGCGCTGAGCTGCTC 653

## RESULT 13

US-09-386-642-9  
 ; Sequence 9, Application US/09386642  
 ; Patent No. 6420157  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Darrow, Andrew  
 ; APPLICANT: Qi, Jensen  
 ; APPLICANT: Andrade-Gordon, Patricia  
 ; TITLE OF INVENTION: Zymogen Activation System  
 ; FILE REFERENCE: ORT-1028  
 ; CURRENT APPLICATION NUMBER: US/09/386,642  
 ; FILING DATE: 1999-08-31  
 ; NUMBER OF SEQ ID NOS: 60  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 9  
 ; LENGTH: 1049  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene  
 ; OTHER INFORMATION: with homo sapien serine protease catalytic domain  
 ; US-09-386-642-9

Query Match 11.0%; Score 159.6; DB 4; Length 1049;  
 Best Local Similarity 54.4%; Pred. No. 6,9e-32;  
 Matches 371; Conservative 0; Mismatches 299; Indels 12; Gaps 2;

QY 247 TCGGAGCGCTGCGAGCTCTGCTTCAAGCGCTCTCTTCACTGCGGAGTCTCTG 306  
 DB 199 TCGGAGCGCTTGGGAGCGGCTTGTTCAGGGCGACGACTCTGTGGCGGTGCTCT 258  
 QY 307 GTGAGCAGAGTGGGCTGTGACGCGCGGCGCACTGCGGAAACAAAGCACTGTGGCTGA 366  
 DB 259 GTAGGTGCGAAGCTGCTTCTTACAGTGGCGCACTGTAAACAAACAAATACAGTACGC 318  
 QY 367 GTAGGAGATGATCACTGTCTCTTCAAGGCGAGAGCTCGCGGAGAGTCTGCTCT 426  
 DB 319 CTGGGAGACCAAGCTTACAGATTAAGATGAGCCCGAGAGCA-----GAAATCTGTGAT 374  
 QY 427 GTTGTCCATCCCAAGTACAGAGGCTGAGGCGCCCATCTGCGCAAGGAGAGAGATGAG 486  
 DB 375 TCAATCATTC-----CACTGTCTTCAAGAGAGAGAGATGAGAGACCAAC 426  
 QY 487 CAGATCTCATGTGTGTAAGCTGCGCAGGCGCGCTGTATGTCGCGGAGCGCTGCGAGCC 546  
 DB 427 CATGATCTGATGCTTCTTCACTGCGGTACAGGCAATCTGCGGATCCAAAGTGAAGCC 486  
 QY 547 CTGAGCTTCCCTACCGCTGTGTCTCAGCGCGGAGACAGATGCTCAGTGTCTGCGGAGC 606  
 DB 487 ATAGCTTGGAGATATGACCAAGCTGTGCGAGAGTGAACCGTCTCAGGCTGTGGGCG 546  
 QY 607 ACCAGGCGCGCGGAGAGTGAAGTAAACAAAGGCGCTGAGCTGCTCAGACTCATATC 666

Db 547 ACTGTCCAGGATCCCCAGAGATTTTCTGACACTCTCACTGTGCAAGTAAATC 606  
QY 667 CTGAGCCCTTAAGAGTGTGAGTCTTCTACCTGCGTGTGCAACAACATGATATGT 726  
Db 607 TTTCCTCCAGAGAGTGTGAGATGCTTACCCGGGAGATCACAAGTGCATGTCTGT 666  
QY 727 GCTGAGCTGAGACCGGGGCGACAGACCTTTGCCAGAGTACTCTGAGAGCCCTGTGTGT 786  
Db 667 GCAGGCGAGCAAGGGGCTGACACGCTGCCAGGGCGATTTCTGAGAGCCCTGTGTGT 726  
QY 787 GACGAGACCTTCCAGGACATCTCTGTTGGGGTGTATTACCCCTGTGTGTGTGTGTGT 846  
Db 727 GATGTGTACTCTCCAGGACATCTCTGAGGGCTGCAAGCCCTGTGTGTGTGTGTGT 786  
QY 847 CCACTGTCTACACCCAGATCTGCAATATGATCTGTGATCAATAAAGTATAGCTTC 906  
Db 787 CTTGGCTCTATACCAACATCTGCGGCTTCTGAGATGTGATCAAGAGATCAGGAGC 846  
QY 907 AACTGATCCAGATGCTACGCTC 928  
Db 847 AAGGGCTGTAGACATGACATC 868

## RESULT 14

US-09-008-271A-19

; Sequence 19, Application US/09008271A

; Patent No. 6203979

## GENERAL INFORMATION:

APPLICANT: Bandman, Olga

Hillman, Jennifer L.

Yue, Henry

Guegler, Karl J.

Corley, Neil C.

Yang, Tom Y.

Shah, Parvi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/008,271A

FILING DATE: 16-Jan-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: &lt;Unknown&gt;

FILING DATE: &lt;Unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: Mohan-Peterson, Sheela

REGISTRATION NUMBER: 41,201

REFERENCE/DOCKET NUMBER: PF-0458 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 994 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: COLNOT27

CLONE: 1798496

SEQUENCE DESCRIPTION: SEQ ID NO: 19 :

US-09-008-271A-19

Query Match 10.7%; Score 156; DB 3; Length 994;  
Best Local Similarity 54.1%; Pred. No. 5,7e-31;  
Matches 368; Conservative 0; Mismatches 300; Indels 12; Gaps 2;

QY 234 GTGCGGCGCGGCTGCGAGCCCTGCGAGGCTGTGCTCTTCAACGCGCTCTGTCACAG 293  
Db 268 GTGCCAACCCCTTGTGCACTTGTGCGAGGCGGCTTGTGCGAGGCGGCGAGCACTTGTG 327  
QY 294 CCGGGGTCTGTGTGTGAGCAAGATTGGGTGTGAGCGCGCGCATCTGCGAACAAGCC 353  
Db 328 TGCGGTGTCTTGT 387  
QY 354 ACTGTGGCTGT 413  
Db 388 ATACAGATGATCTGT 443  
QY 414 GACGATCTGT 473  
Db 444 AATACCTGT 495  
QY 474 GCGAAGGATGAGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 533  
Db 496 GAGGACCAACAACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 555  
QY 534 CCGCGTCCGCGGCTGCGAGCTTCCCTACCGCTGTGTGTGTGTGTGTGTGTGTGTGT 593  
Db 556 CAAGTGAAGCCCTTGT 615  
QY 594 TGT 653  
Db 616 CTCAGGCTGT 675  
QY 654 CAGCATCACTATCTGT 713  
Db 676 AGAATTAATAATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 735  
QY 714 CAACATGATATGT 773  
Db 736 TGCGATGT 795  
QY 774 CCGCTGT 833  
Db 796 CCGCTGT 855  
QY 834 CTGCGCCAGCATCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 893  
Db 856 GAGTCCGACAACTGT 915  
QY 894 AGTCATAGCTTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 913  
Db 916 GATCATAGGCAAGGGCT 935

## RESULT 15

US-09-070-526-1

; Sequence 1, Application US/09070526

; Patent No. 6100059

## GENERAL INFORMATION:

APPLICANT: SOUTHAN, CHRISTOPHER

BORGESS, NICOLA

TITLE OF INVENTION: No. 6100059e1 Compounds

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: RATNER &amp; PRESTIA

STREET: P.O. BOX 980

CITY: VALLEY FORGE

STATE: PA

COUNTRY: USA

ZIP: 19482

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 04:52:03 ; Search time 425 seconds

(without alignments)  
9235.254 Million cell updates/sec

Title: US-10-021-368-2

Perfect score: 1454  
Sequence: 1 ACCAGGGGACAGACACAGGC.....GTGCTGAAAAA 1454

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: N\_Geneseq\_19Jun03.\*  
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1454	100.0	1454	18	AAT44111 Human NES1 CDNA.
2	1454	100.0	1454	24	AB576452 CDNA encoding huma
3	1454	100.0	1454	24	AB106519 Human serine prote
4	1442	99.2	1457	24	AB234966 Human gene express
5	829.4	57.0	831	24	ABK31771 DNA encoding novel
6	627	43.1	750	24	ABO5252 Human ovarian anti
7	508.4	35.0	532	24	AB163341 Breast cancer rela
8	508.4	35.0	532	24	AB163744 Breast cancer rela

9	508.4	35.0	532	24	AB167277	Thyroid cancer rel
10	508.4	35.0	532	24	AB182556	Human ovarian can
11	448.4	30.8	454	24	AB182657	Human ovarian can
12	447.4	30.8	449	24	AB182374	Human ovarian can
13	445.4	30.6	470	24	AB179800	Human ovarian can
14	438.4	30.2	468	24	AB179803	Human ovarian can
15	383.8	26.4	399	24	AB182197	Human ovarian can
16	379.6	26.1	430	24	AB181770	Human ovarian can
17	376	25.9	336	24	AB194495	Gene #93 used to
18	376	25.9	396	24	AB163950	Breast cancer rela
19	376	25.9	396	24	AB164839	Lung cancer relate
20	376	25.9	396	24	AB166152	Lung cancer relate
21	376	25.9	396	24	AB181574	Human ovarian can
22	367	25.2	434	24	AB179853	Human ovarian can
23	353	24.3	373	22	AA129327	Human ovarian can
24	353	24.3	373	25	AB233513	Human ovarian can
25	352	24.2	421	24	AB181769	Human ovarian can
26	345.4	23.8	359	24	AB182352	Human ovarian can
27	335.6	23.1	377	21	AA280588	Human ovarian can
28	329.4	22.7	349	24	AB180212	Human ovarian can
29	319	21.9	321	22	AA129411	Human ovarian can
30	319	21.9	321	25	AB233597	Human ovarian can
31	295	20.3	339	24	AB182894	Human ovarian can
32	274.6	18.9	292	24	AB181595	Human ovarian can
33	267.2	18.4	283	24	AB167055	Thyroid cancer rel
34	267.2	18.4	283	24	AB167485	Thyroid cancer rel
35	267.2	18.4	283	24	AB182297	Human ovarian can
36	256	17.6	274	24	AB179814	Human ovarian can
37	243.4	16.7	267	24	AB179810	Human ovarian can
38	203.4	14.0	205	22	AA524516	Human ovarian can
39	203.4	14.0	319	22	AA583124	Human ovarian can
40	173.8	12.0	1091	21	AA581114	Human ovarian can
41	173.8	12.0	1091	21	AA776571	Human PRO1303 cDNA
42	173.8	12.0	1091	21	AA337075	Human PRO1303 (UNG
43	173.8	12.0	1091	22	AA54341	DNA encoding prote
44	173.8	12.0	1185	22	AA198653	Human EST-derived
45	165.8	11.4	1166	22	AA14841	Human PS13 consen

#### ALIGNMENTS

RESULT 1	
AA144111	
ID	AA144111 standard; cDNA; 1454 BP.
XX	AC
XX	AA144111;
DT	28-FEB-1997 (first entry)
XX	XX
DE	Human NES1 CDNA.
XX	XX
KW	NES1; malignancy; cancer; breast carcinoma; cervix carcinoma;
KW	prostate carcinoma; gene therapy; diagnosis; prognosis;
KW	serine protease; ss.
XX	XX
OS	Homo sapiens.
XX	XX
PH	Key
FT	CDS
FT	Location/Qualifiers
FT	82..912
FT	/*tag= a
FT	1348..1353
FT	polYA_signal
XX	/*tag= b
XX	XX
PN	W09639175-A1.
XX	XX
PD	12-DEC-1996.
XX	XX
PF	21-MAY-1996; 96WC-US07343.
XX	XX
PR	06-JUN-1995; 95US-0467155.
XX	XX
PA	(NEW-) NEW ENGLAND MEDICAL CENT HOSPITALS INC.











Db 541 CGGGCCCTGAGCTTCCCTACCGCTGCTCAGCCCGAGACAGTCCAGGTTGCTGCG 600  
 QY TGGGGACACACGCGCGCCCGAGAGGTGAAATGACAAAGAGGCGCTGACCTGCTCCAGATC 660  
 Db 601 TGGGGACACACGCGCGCCCGAGAGGTGAAATGACAAAGAGGCGCTGACCTGCTCCAGATC 660  
 QY 601 TGGGGACACACGCGCGCCCGAGAGGTGAAATGACAAAGAGGCGCTGACCTGCTCCAGATC 660  
 Db 661 ACTATCTGAGCCCTTAAAGAGGTGAGGTCTTCTTACCTGCGGTGCTCACCACAAACATG 720  
 QY 661 ACTATCTGAGCCCTTAAAGAGGTGAGGTCTTCTTACCTGCGGTGCTCACCACAAACATG 720  
 Db 721 ATATGCTGTGACCTGAGACCGGGGCGCAGAGCCCTTGGCCAGATGACTGTGAGAGCCCTG 780  
 QY 721 ATATGCTGTGACCTGAGACCGGGGCGCAGAGCCCTTGGCCAGATGACTGTGAGAGCCCTG 780  
 Db 781 GTCTGTGACGAGACCCCTTCCAAAGGATCTCTGCGGGGTGTTTACCCCTGAGGCTGCGC 840  
 QY 781 GTCTGTGACGAGACCCCTTCCAAAGGATCTCTGCGGGGTGTTTACCCCTGAGGCTGCGC 840  
 Db 841 CAGCATCCAGTGTCTTACACCCAGATCTGCATAATATGCTCTGATCAATTAAGTCATA 900  
 QY 841 CAGCATCCAGTGTCTTACACCCAGATCTGCATAATATGCTCTGATCAATTAAGTCATA 900  
 Db 901 CGCTCCAACTGATCCAGATGCTACGCTCCAGCTGATCCAGATGCTATGCTCTGCTGATC 960  
 QY 901 CGCTCCAACTGATCCAGATGCTACGCTCCAGCTGATCCAGATGCTATGCTCTGCTGATC 960  
 Db 961 CAGATGCCAGAGGCTCATGCTCATCTTCTTCTCCAGTGGCTGAGACTCTCCCT 1020  
 QY 961 CAGATGCCAGAGGCTCATGCTCATCTTCTTCTCCAGTGGCTGAGACTCTCCCT 1020  
 Db 1021 TGTCTGACCTGTTTAAACCTCTGCGCGCTCCAGCTCAACCTTAAACCTTCCCTCTACCT 1080  
 QY 1021 TGTCTGACCTGTTTAAACCTCTGCGCGCTCCAGCTCAACCTTAAACCTTCCCTCTACCT 1080  
 Db 1081 CATTCGCCCACTTCCCTCTCTGCTGCTGATGAGCTGAAATGACAGAAAGTGGTGG 1140  
 QY 1081 CATTCGCCCACTTCCCTCTCTGCTGCTGATGAGCTGAAATGACAGAAAGTGGTGG 1140  
 Db 1141 CAAAGGTTTATTCAGAGAAAGCCAGAAAGCCGCTGATCACCACCTCTGAGACAGTTA 1200  
 QY 1141 CAAAGGTTTATTCAGAGAAAGCCAGAAAGCCGCTGATCACCACCTCTGAGACAGTTA 1200  
 Db 1201 CTGGGGTCCACCAACCTGATCTCTGCGCACTCCCGCTGCTGACTTGGGCAAGCA 1260  
 QY 1201 CTGGGGTCCACCAACCTGATCTCTGCGCACTCCCGCTGCTGACTTGGGCAAGCA 1260  
 Db 1260 AGTCCCTCTCTGAACTTCTCTCATCTGCAAAATGGAACATGACGTGCCCTAC 1319  
 QY 1260 AGTCCCTCTCTGAACTTCTCTCATCTGCAAAATGGAACATGACGTGCCCTAC 1319  
 Db 1321 TCTTGAACATGTTGTGAGAGACATATATTAATGATGTAATCTTCAATGCAAT 1380  
 QY 1321 TCTTGAACATGTTGTGAGAGACATATATTAATGATGTAATCTTCAATGCAAT 1380  
 Db 1381 GTCATGTAAAGCTTAAACAGTGGGTGAGTCTGACTTAAAGTTAACTGTTGCGTG 1440  
 QY 1381 GTCATGTAAAGCTTAAACAGTGGGTGAGTCTGACTTAAAGTTAACTGTTGCGTG 1440  
 Db 1441 AAAAAGGCTTAAACAGTGGGTGAGTCTGACTTAAAGTTAACTGTTGCGTG 1439  
 QY 1441 AAAAAGGCTTAAACAGTGGGTGAGTCTGACTTAAAGTTAACTGTTGCGTG 1439  
 Db 1440 AAAAAAAAAAAAAA 1453  
 QY 1440 AAAAAAAAAAAAAA 1453

## RESULT 5

ID ABK31771 standard, DNA: 831 BP.

AC ABK31771;

XX 23-APR-2002 (first entry)

XX DNA encoding novel human protease #28.

XX

KW Human; protease; cancer; immune-related disorder; cardiovascular disease;  
 KW neuronal-associated disease; metabolic disorder; inflammatory disorder;  
 KW nervous system disorder; sexual dysfunction; pain; mood disorder;  
 KW hypertension; psychotic disorder; neurological disorder; dyskinesia;  
 KW viral infection; human immunodeficiency virus; HIV; non-viral infection;  
 KW ocular disease; cytostatic; gene; ds.  
 OS Homo sapiens.  
 PN NC0200200860-A2.  
 PD 03-JAN-2002.  
 PF 26-JUN-2001; 2001MO-US20171.  
 PR 26-JUN-2000; 2000US-214047P.  
 PA (SUGR-) SUGEN INC.  
 PI Plowman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;  
 PI Charyczak G;  
 DR WPI, 2002-139913/18.  
 DR P-PSDB; AA082729.  
 PT Nucleic acids encoding novel human proteases, useful for useful for  
 PT treating diseases and disorders such as cancers, immune-related  
 PT diseases and disorders, cardiovascular disease (e.g. restenosis) and  
 PT inflammatory disorders -  
 XX  
 XX Claim 26, Fig 1CC; 313pp; English.  
 CC The present invention relates to the isolation of novel human  
 CC proteases, and the nucleic acids encoding them. The sequences of  
 CC the invention are useful for treating diseases and disorders such as  
 CC cancers (e.g. breast, colon, lung), immune-related diseases and disorders  
 CC (e.g. inflammatory diseases and asthma), cardiovascular diseases  
 CC (e.g. restenosis and coronary thrombosis), brain or neuronal-associated  
 CC diseases, metabolic disorders (e.g. diabetes, obesity), inflammatory  
 CC disorders (e.g. rheumatoid arthritis and psoriasis), central or  
 CC peripheral nervous system diseases, migraines, pain, sexual dysfunction,  
 CC mood disorders, attention disorders, cognitive disorders, hypotension,  
 CC hypertension, psychotic disorders, neurological disorders  
 CC (e.g. Alzheimer's disease, Parkinson's disease) and dyskinesias.  
 CC The nucleic acids and polypeptides are also useful for treating viral  
 CC infections caused by human immunodeficiency virus (HIV), and non-viral  
 CC infections such as ocular disease (e.g. glaucoma) and macular  
 CC degeneration. ABK31744-ABK31802 represent DNA sequences encoding for  
 CC the novel human proteases of the invention.  
 XX  
 SQ Sequence 831 BP; 146 A; 287 C; 245 G; 153 T; 0 other;

Query Match 57.0%; Score 829.4; DB 24; Length 831;  
 Best Local Similarity 99.9%; Pred. No. 7.9e-179;  
 Matches 830; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 82 ATGAGAGCTCCGACCTCCACCTTCCGCGGCTCTGAGGCGCGGAGCTTGGCGAGCTG 141  
 Db 1 ATGAGAGCTCCGACCTCCACCTTCCGCGGCTCTGAGGCGCGGAGCTTGGCGAGCTG 60  
 QY 142 CTGGCGGCTGATGAGCGCACTTGGGCGCGAGAGGCGGCGCTCTCCCAAAACGAC 201  
 Db 61 CTGGCGGCTGATGAGCGCACTTGGGCGCGAGAGGCGGCGCTCTCCCAAAACGAC 120  
 QY 202 ACGCGCTTGAACCCGGAAGCTATGCGCGCGCGCGCGCGCTCGAGAGCTTGGCGAG 261  
 Db 121 ACGCGCTTGAACCCGGAAGCTATGCGCGCGCGCGCGCGCTCGAGAGCTTGGCGAG 180  
 QY 262 GTCTGCTCTTAAAGGCTCTCTGTTCACTGCGCGGCTGCTGAGGACCAAGATTGG 321  
 Db 181 GTCTGCTCTTAAAGGCTCTCTGTTCACTGCGCGGCTGCTGAGGACCAAGATTGG 240  
 QY 322 GTCTGAGCGCGCGCACTGCGGAAACAAAGCACTGTGGGCTCGAGTAGGGGATATCAC 381

DQ 241 GTGCTGACGGCCGCGACTGCGGAAACAGGCCACTGTGGGCTCGAGTAGGGGATGACCAAC 300  
 QY 382 CTGCTGCTCTTTCAGGGCCGAGAGCTCCGCGGAGACACTCGCTGCTGCTCCATCCCAAG 441  
 DB 301 CTGCTGCTCTTTCAGGGCCGAGAGCTCCGCGGAGACACTCGCTGCTGCTCCATCCCAAG 360  
 QY 442 TACCACCAAGGAGCTCAGGCGCCCATCTGCCAAGGCGAAGGATGAGACCAATCTCATGTTG 501  
 DB 361 TACCACCAAGGAGCTCAGGCGCCCATCTGCCAAGGCGAAGGATGAGACCAATCTCATGTTG 420  
 QY 502 CTAAAGCTGGCCGAGCCGCTGAGTGGCGGGGCGCCCGGCTCGGGGCGCTGAGCTTCCCTAC 561  
 DB 421 CTAAAGCTGGCCGAGCCGCTGAGTGGCGGGGCGCCCGGCTCGGGGCGCTGAGCTTCCCTAC 480  
 QY 562 CGCTGTGCTCAGCCCGGAGACCAAGTGCAGGTTGCTGCTGAGGCGACCAAGCGCCCGCG 621  
 DB 481 CGCTGTGCTCAGCCCGGAGACCAAGTGCAGGTTGCTGCTGAGGCGACCAAGCGCCCGCG 540  
 QY 622 AGAGTGAAGTCAACAAAGGCGCTGACCTGCTCCAGATCACTATCTGAGCCCTTAAAGAG 681  
 DB 541 AGAGTGAAGTCAACAAAGGCGCTGACCTGCTCCAGATCACTATCTGAGCCCTTAAAGAG 600  
 QY 682 TGTGAGGCTCTTACCTGCGGCTGCTGACCAACCAATGATATGCTGAGCTGAGACCGG 741  
 DB 601 TGTGAGGCTCTTACCTGCGGCTGCTGACCAACCAATGATATGCTGAGCTGAGACCGG 660  
 QY 742 GGGCAGAGACCTCTGCGGAGATGCTGAGAGGCGCCCTGCTGAGGAGACCGCTCCAA 801  
 DB 661 GGGCAGAGACCTCTGCGGAGATGCTGAGAGGCGCCCTGCTGAGGAGACCGCTCCAA 720  
 QY 802 GGGATCTCTCTGAGGAGGTTTACCCCTGCTGAGCTGCTGAGGAGACCGCTGCTGAGAC 861  
 DB 721 GGGATCTCTCTGAGGAGGTTTACCCCTGCTGAGCTGCTGAGGAGACCGCTGCTGAGAC 780  
 QY 862 CAGATCTGCAAAATACATGCTGATCAATAAAGTATATGCTGCAACTGA 912  
 DB 781 CAGATCTGCAAAATACATGCTGATCAATAAAGTATATGCTGCAACTGA 831

RESULT 6  
 ABQ55252  
 ID ABQ55252 standard; cDNA, 750 BP.  
 AC ABQ55252;  
 XX  
 DT 22-AUG-2002 (first entry)  
 XX  
 DE Human ovarian antigen HCOPE27 cDNA, SEQ ID NO:1132.  
 XX  
 KM Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KM ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KM infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KM PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KM inflammatory condition; immune disorder; blood disorder;  
 KM cardiovascular disorder; respiratory disorder; neurological disorder;  
 KM gastrointestinal disorder; urinary system disorder; drug screening;  
 KM gene therapy; chromosome mapping; forensic analysis;  
 KM antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KM antiinflammatory; gynaecological; reproductive; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200677-A1.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 07-JUN-2001; 2001WO-US18569.  
 XX  
 PR 07-JUN-2000; 2000US-209467P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.

P1 Birse CE, Rosen CA;  
 XX  
 DR WPI; 2002-147878/19.  
 DR P-FSDB; ABP42175.  
 XX  
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
 PT ovarian cancer), immune disorders, cardiovascular disorders and  
 PT neurological diseases -  
 XX  
 PS Claim 1; SEQ ID NO 1132; 2922pp; English.  
 XX  
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ4131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovary and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., congenital and acquired  
 CC vaginitis), immune disorders (e.g., autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC invention represents cDNA encoding a human ovarian antigen of the  
 CC sequence:  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WPI  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 SQ Sequence 750 BP; 144 A; 254 C; 173 G; 169 T; 10 other;  
 XX  
 Query Match 43.1%; Score 627; DB 24; Length 750;  
 Best Local Similarity 98.2%; Pred. No. 7.4e-133;  
 Matches 652; Conservative 0; Mismatches 10; Indels 2; Gaps 2;  
 QY 510 GGGCAGGCGCTGATGTCGCGGAGCGCCGCTGCGGAGCTTCCCTACCGCTGTGC 569  
 DB 1 GGCAGGCGCTGATGTCGCGGAGCGCCGCTGCGGAGCTTCCCTACCGCTGTGC 60  
 QY 570 TCAAGCCCGAGAGCAATGCGAGTTCCTGCTGAGGAGACCAAGCGCCCGGAGATGA 629  
 DB 61 TCAAGCCCGAGAGCAATGCGAGTTCCTGCTGAGGAGACCAAGCGCCCGGAGATGA 120  
 QY 630 GTACAACAAGGCGCTGACCTGCTCCAGATCACTATCTGAGCCCTTAAAGATGAGGT 689  
 DB 121 GTACAACAAGGCGCTGACCTGCTCCAGATCACTATCTGAGCCCTTAAAGATGAGGT 180  
 QY 690 CTCTTACCTGCGGCTGCTGACCAACAATGATGCTGAGCTGAGACCGGCGCAGGA 749  
 DB 181 CTCTTACCTGCGGCTGCTGACCAACAATGATGCTGAGCTGAGACCGGCGCAGGA 240  
 QY 750 CCCTTGCAGAGTATCTGAGAGCGCCCTGCTGCTGAGAGACCTTCAAGGATCT 809  
 DB 241 CCCTTGCAGAGTATCTGAGAGCGCCCTGCTGCTGAGAGACCTTCAAGGATCT 300  
 QY 810 CTCTGAGGAGTTTACCTGCTGCTGCGACATCCAGCTGCTTACACCAATCTG 869  
 DB 301 CTCTGAGGAGTTTACCTGCTGCTGCGACATCCAGCTGCTTACACCAATCTG 360

Qy	870	CAAAATGATATGCTGGATCATTAAGTCATACGCTCAAGTCAGATGCTACGCTGC	9239
Dd	361	CAAAATGATATGCTGGATCATTAAGTCATACGCTCAAGTCAGATGCTACGCTGC	420
Qy	930	AGCTGATCCAGATGTTATGCTCTGCTGATCCAGATGCCAGAGCTCCATGTCATCC	9899
Dd	421	AGCTGATCCAGATGTTATGCTCTGCTGATCCAGATGCCAGAGCTCCATGTCATCC	480
Qy	990	TCTTCCCTCCAGTCGGCTGAATCTCCGCTGTGCTGCAGCTGTTCAAACTCTG-CCGCC	1040
Dd	481	TCTTCCCTCCAGTCGGCTGAATCTCCGCTGTGCTGCAGCTGTTCAAACTCTGCGCC	540
Qy	1049	CTCCACACCTCTTAAACATCTCCCTCTCACTCATTTCCCACTATCCCATTTCTGTC	110
Dd	541	CTCCACACCTCTTAAACATCTCCCTCTCACTCATTTCCCACTATCCCATTTCTGTC	600
Qy	1109	CTGTACTGGAAGCTGAATGAGAGAGTGGCAAGGTTATTTCAGAGAGCCAGAAA	116
Dd	601	CTGTACTGGAAGCTTAAATGCAAGAGTGGCCAGG-TTATTTCAGAGAGCCAGAAA	6599
Qy	1169	GCCG 1172	
Dd	660	GCGG 663	
RESULT 7			
ABL63341			
ID	ABL63341	standard; DNA; 532 BP.	
XX	ABL63341;		
XX	15-MAY-2002	(first entry)	
DE	Breast cancer related gene sequence SEQ ID NO:1678.		
KM	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;		
KM	stomach; lung; prostate; pancreas; carcinoma; antileucmur; cancerous;		
KM	cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;		
XX	gene; ds.		
XX	Homo sapiens.		
XX	MOJ00194629-A2.		
PD	13-DEC-2001.		
XX	30-MAY-2001;	2001MO-US10838.	
XX	05-JUN-2000;	2000US-209473P.	
PR	05-JUN-2000;	2000US-209531P.	
PR	18-SEP-2000;	2000US-233133P.	
PR	18-SEP-2000;	2000US-233617P.	
PR	20-SEP-2000;	2000US-234009P.	
PR	20-SEP-2000;	2000US-234034P.	
PR	20-SEP-2000;	2000US-234052P.	
PR	22-SEP-2000;	2000US-234509P.	
PR	22-SEP-2000;	2000US-234567P.	
PR	25-SEP-2000;	2000US-234923P.	
PR	25-SEP-2000;	2000US-234924P.	
PR	25-SEP-2000;	2000US-235077P.	
PR	25-SEP-2000;	2000US-235087P.	
PR	25-SEP-2000;	2000US-235134P.	
PR	25-SEP-2000;	2000US-235280P.	
PR	26-SEP-2000;	2000US-235631P.	
PR	26-SEP-2000;	2000US-235638P.	
PR	27-SEP-2000;	2000US-235711P.	
PR	27-SEP-2000;	2000US-235720P.	
PR	27-SEP-2000;	2000US-235840P.	
PR	27-SEP-2000;	2000US-235863P.	
PR	28-SEP-2000;	2000US-236028P.	
PR	28-SEP-2000;	2000US-236032P.	
PR	28-SEP-2000;	2000US-236033P.	
PR	28-SEP-2000;	2000US-236034P.	

PR	28-SEP-2000;	2000US-236109P.	
PR	28-SEP-2000;	2000US-236111P.	
PR	29-SEP-2000;	2000US-236847P.	
PR	29-SEP-2000;	2000US-236891P.	
PR	02-OCT-2000;	2000US-237172P.	
PR	02-OCT-2000;	2000US-237173P.	
PR	02-OCT-2000;	2000US-237278P.	
PR	02-OCT-2000;	2000US-237294P.	
PR	02-OCT-2000;	2000US-237315P.	
PR	03-OCT-2000;	2000US-237425P.	
PR	03-OCT-2000;	2000US-237598P.	
PR	03-OCT-2000;	2000US-237604P.	
PR	03-OCT-2000;	2000US-237606P.	
PR	03-OCT-2000;	2000US-237608P.	
PR	01-NOV-2000;	2000US-244867P.	
PR	01-NOV-2000;	2000US-245084P.	
PA	(AVAL-)	AVALON PHARM.	
XX			
XX	Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;		
P1	Soppet DR, Weaver Z;		
XX			
DR	WPI; 2002-188264/24.		
XX			
PT	Screening for anti-neoplastic agent involves exposing cells to a		
PT	chemical agent to be tested for anti-neoplastic activity, and		
PT	determining a change in expression of a gene of a signature gene set		
PS	Claim 1; SEQ ID 1678; 44pp; English.		
XX			
CC	The present invention describes a method (M1) for screening for an		
CC	anti-neoplastic agent. The method involves exposing cells to a chemical		
CC	agent to be tested for anti-neoplastic activity, determining a change in		
CC	expression of at least one gene (I) of a signature gene set, where (I) has		
CC	comprises a sequence (S) selected from 847 sequences (given in Abn61664		
CC	to Abn70110), or is at least 95% identical to (S), where a change in		
CC	expression is indicative of anti-neoplastic activity. (I) has cytostatic		
CC	activity and can be used in gene therapy. M1 can be used for screening		
CC	an anti-neoplastic agent, and can be used for producing a product which		
CC	is the data collected with respect to the anti-neoplastic agent as a		
CC	result of M1, and the data is sufficient to convey the chemical		
CC	structure and/or properties of the agent. M1 can be used in the		
CC	treatment of cancer such as colon, breast, stomach, lung, thyroid,		
CC	oesophagaeal, ovarian, kidney, prostate or pancreatic cancer,		
CC	adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,		
CC	infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine		
CC	carcinoma, papillary carcinoma and Wilms tumour.		
XX			
SO	Sequence 532 BP; 110 A; 163 C; 113 G; 126 T; 0 other;		
Query Match	35.0%; Score 508.4; DB 24; Length 532;		
Best Local Similarity	99.4%; Pred. No. 6e-106;		
Matches 531; Conservative	0; Mismatches 1; Indels 2; Gaps 2;		
QY	712 AACACATGATATGTGCTGAGCTGACCGGGGGCCAGACCCCTTCCAGAGTACTGGA	771	
DB	1 AACACATGATATGTGCTGAGCTGACCGGGGGCCAGACCCCTTCCAGAGTACTGGA	60	
QY	772 GGGCCCTGGTCTGTGTGACGAGACCTCTCCAGAGATCTCTGTGGGGTGTATTACCTGT	831	
DB	61 GGGCCCTGGTCTGTGTGACGAGACCTCTCCAGAGATCTCTGTGGGGTGTATTACCTGT	120	
QY	832 GGCTCTGGCCGAGTCCAGCTGTCTACACCCAGATCTGCAGATATCATGTCTCGATCAAT	891	
DB	121 GGCTCTGTGCGAGATCCAGCTGTCTACACCCAGATCTGCAGATATCATGTCTCGATCAAT	179	
QY	892 AAAGTCATACCTCTCCAACTGATCCAGATGCTACGCTCCAGCTATCCAGATGTTATGCTC	951	
DB	180 AAAGTCATACCTCTCCAACTGATCCAGATGCTACGCTCCAGCTATCCAGATGTTATGCTC	239	
QY	952 CTGTGATATCCAGATGCCAGAGGCTCCATGCTGCATCTCTTCTCCCAAGTCCGCTGAA	1011	

Db 240 CTGCTGATCCAGATGCCAGAGGCTCCATGCTCTTCTCCAGTCGCTGAA 299  
 QY 1012 CTCTCCCTGCTGCTGCACTGTTTAAACCTTGCGCCCTCCACACCTTAAACATCTCC 1071  
 Db 300 CTCTCCCTGCTGCTGCACTGTTTAAACCTTGCGCCCTCCACACCTTAAACATCTCC 359  
 QY 1072 CTCTACCTCATTCCTCCACCTATCCCATCTCTCTGCTTACTGAACTGAATGCAAG 1131  
 Db 360 CTCTACCTCATTCCTCCACCTATCCCATCTCTCTGCTTACTGAACTGAATGCAAG 419  
 QY 1132 AAGGTGGTGAAGGTTTATTCAGAGAGCCGAGGCGGTCAACCCAGCTCTGA 1191  
 Db 420 AAGGTGGTGAAGGTTTATTCAGAGAGCCGAGGCGGTCAACCCAGCTCTGA 479  
 QY 1192 GAGCAGTCTACTGAGGTCACCCCACTGACTTCTCTGCTCCACTCCCGCTGTGTG 1245  
 Db 480 GAGCAGTCTACTGAGGTCACCCCACTGACTTCTCTGCTCCACTCCCGCTGTGTG 532

RESULT 8  
 ABL63744  
 ID ABL63744 standard; DNA; 532 BP.  
 XX  
 AC ABL63744;  
 XX  
 DT 15-MAY-2002 (first entry)  
 XX  
 DE Breast cancer related gene sequence SEQ ID NO:2081.  
 XX  
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lungs; prostate; pancreas; carcinoma; adenocarcinoma;  
 KW cytoskeletal; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FN W0200194629-A2.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PF 30-MAY-2001; 2001MO-US10838.  
 XX  
 PR 05-JUN-2000; 2000US-209473P.  
 PR 05-JUN-2000; 2000US-209531P.  
 PR 18-SEP-2000; 2000US-233133P.  
 PR 18-SEP-2000; 2000US-233617P.  
 PR 20-SEP-2000; 2000US-234009P.  
 PR 20-SEP-2000; 2000US-234034P.  
 PR 20-SEP-2000; 2000US-234052P.  
 PR 22-SEP-2000; 2000US-234509P.  
 PR 22-SEP-2000; 2000US-234567P.  
 PR 25-SEP-2000; 2000US-234923P.  
 PR 25-SEP-2000; 2000US-234924P.  
 PR 25-SEP-2000; 2000US-235077P.  
 PR 25-SEP-2000; 2000US-235082P.  
 PR 25-SEP-2000; 2000US-235134P.  
 PR 25-SEP-2000; 2000US-235280P.  
 PR 26-SEP-2000; 2000US-235637P.  
 PR 27-SEP-2000; 2000US-235638P.  
 PR 27-SEP-2000; 2000US-235711P.  
 PR 27-SEP-2000; 2000US-235720P.  
 PR 27-SEP-2000; 2000US-235840P.  
 PR 27-SEP-2000; 2000US-235863P.  
 PR 28-SEP-2000; 2000US-236028P.  
 PR 28-SEP-2000; 2000US-236032P.  
 PR 28-SEP-2000; 2000US-236033P.  
 PR 28-SEP-2000; 2000US-236034P.  
 PR 28-SEP-2000; 2000US-236109P.  
 PR 28-SEP-2000; 2000US-236111P.  
 PR 29-SEP-2000; 2000US-236842P.  
 PR 29-SEP-2000; 2000US-236891P.  
 PR 02-OCT-2000; 2000US-237172P.  
 PR 02-OCT-2000; 2000US-237173P.

PR 02-OCT-2000; 2000US-237278P.  
 PR 02-OCT-2000; 2000US-237294P.  
 PR 02-OCT-2000; 2000US-237295P.  
 PR 02-OCT-2000; 2000US-237316P.  
 PR 03-OCT-2000; 2000US-237325P.  
 PR 03-OCT-2000; 2000US-237598P.  
 PR 03-OCT-2000; 2000US-237604P.  
 PR 03-OCT-2000; 2000US-237606P.  
 PR 03-OCT-2000; 2000US-237608P.  
 PR 01-NOV-2000; 2000US-244867P.  
 PR 01-NOV-2000; 2000US-245084P.  
 PA (AVALON PHARM.  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX MPI: 2002-188264/24.  
 XX  
 XX Screening for anti-neoplastic agent involves exposing cells to a  
 PT chemical agent to be tested for anti-neoplastic activity, and  
 PT determining a change in expression of a gene of a signature gene set  
 XX  
 PS Claim 1; SEQ ID 2081; 44pp; English.  
 XX  
 CC The present invention describes a method (M1) for screening for an  
 CC anti-neoplastic agent. The method involves exposing cells to a chemical  
 CC agent to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 CC to ABL70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening  
 CC an anti-neoplastic agent, and can be used for producing a product which  
 CC is the data collected with respect to the anti-neoplastic agent as a  
 CC result of M1, and the data is sufficient to convey the chemical  
 CC structure and/or properties of the agent. M1 can be used in the  
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer;  
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
 CC carcinoma, papillary carcinoma and Wilms' tumour.  
 XX  
 XX Sequence 532 BP; 110 A; 183 C; 113 G; 126 T; 0 other;

Query Match 35.0%; Score 508.4; DB 24; Length 532;  
 Best Local Similarity 99.4%; Pred. No. 66-106;  
 Matches 531; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 712 AACACATGATATGCTGCACTGACCGGGGCCAGAGCCCTTGCCAGAGTACTTGA 771  
 Db 1 AACACATGATATGCTGCACTGACCGGGGCCAGAGCCCTTGCCAGAGTACTTGA 60  
 QY 772 GGGCCCTGCTGCTGAGAGAGACCCCTCCAGAGCATCTCTCTGAGGCTTTAACCCCTG 831  
 Db 61 GGGCCCTGCTGCTGAGAGAGACCCCTCCAGAGCATCTCTCTGAGGCTTTAACCCCTG 120  
 QY 832 GGGCTGCGCAGCATCCAGCTGTCTTACACCCAGATCTGCAATATCATGCTCGATCAAT 891  
 Db 121 GGGCTGCGCAGCATCCAGCTGTCTTACACCCAGATCTGCAATATCATGCTCGATCAAT 179  
 QY 892 AAGGTCAATGCTGCACTGATTCAGATGCTTCAAGCTTCAAGTATCCAGATGTTATGCTC 951  
 Db 180 AAGGTCAATGCTGCACTGATTCAGATGCTTCAAGTATCCAGATGTTATGCTC 239  
 QY 952 CTGCTGATCCAGATGCCAGAGGCTTCATCGTTCATCTCTTCCCTCCAGTGGCTGAA 1011  
 Db 240 CTGCTGATCCAGATGCCAGAGGCTTCATCGTTCATCTCTTCCCTCCAGTGGCTGAA 299  
 QY 1012 CTCTCCCTTGTCTGCACTGTTTAAACCTTGCGCCCTCCACACCTTAAACATCTCC 1071  
 Db 300 CTCTCCCTTGTCTGCACTGTTTAAACCTTGCGCCCTCCACACCTTAAACATCTCC 359

QY 1072 CTCTCACTCATTTCCCTCCCTATTCCTCTGCTGTACTGAAAGCTGAAATGCAGG 1131  
 DB 360 CTCTCACTCATTTCCCTCCCTATTCCTCTGCTGTACTGAAAGCTGAAATGCAGG 419  
 QY 1132 AAGTGTGGCAAGGTTTATTCGAGAGAGCAGAGCCGCTATACCCAGCTCTGA 1191  
 DB 420 AAGTGTGGCAAGGTTTATTCGAGAGAGCAGAGCCGCTATACCCAGCTCTGA 479  
 QY 1192 GAGCAGTTACTGGGGTCAACCCACCTGACTTCTCTGCACTCCCGCTGTGTG 1245  
 DB 480 GAGCAGTTACTGGGGTCA-CCAACTGACTTCTCTGCACTCCCGCTGTGTG 532

## RESULT 9

ABL67277  
 ID ABL67277 standard; DNA; 532 BP.

AC ABL67277;

DT 15-MAY-2002 (first entry)

XX Thyroid cancer related gene sequence SEQ ID NO:5614.

XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW cytosarcoma; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytosarcoma; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;  
 XX gene; ds.

XX Homo sapiens.

XX WO200194629-A2.

XX 13-DEC-2001.

XX 30-MAY-2001; 2001WO-US10838.

XX 05-JUN-2000; 2000US-209473P.  
 PR 05-JUN-2000; 2000US-209531P.  
 PR 18-SEP-2000; 2000US-233133P.  
 PR 18-SEP-2000; 2000US-233617P.  
 PR 20-SEP-2000; 2000US-234003P.  
 PR 20-SEP-2000; 2000US-234003P.  
 PR 20-SEP-2000; 2000US-234052P.  
 PR 22-SEP-2000; 2000US-234567P.  
 PR 22-SEP-2000; 2000US-234567P.  
 PR 25-SEP-2000; 2000US-234923P.  
 PR 25-SEP-2000; 2000US-234923P.  
 PR 25-SEP-2000; 2000US-235077P.  
 PR 25-SEP-2000; 2000US-235082P.  
 PR 25-SEP-2000; 2000US-235134P.  
 PR 25-SEP-2000; 2000US-235280P.  
 PR 26-SEP-2000; 2000US-235637P.  
 PR 26-SEP-2000; 2000US-235638P.  
 PR 27-SEP-2000; 2000US-235711P.  
 PR 27-SEP-2000; 2000US-235720P.  
 PR 27-SEP-2000; 2000US-235840P.  
 PR 27-SEP-2000; 2000US-235863P.  
 PR 28-SEP-2000; 2000US-236028P.  
 PR 28-SEP-2000; 2000US-236032P.  
 PR 28-SEP-2000; 2000US-236033P.  
 PR 28-SEP-2000; 2000US-236034P.  
 PR 28-SEP-2000; 2000US-236109P.  
 PR 28-SEP-2000; 2000US-236111P.  
 PR 29-SEP-2000; 2000US-236842P.  
 PR 29-SEP-2000; 2000US-236891P.  
 PR 02-OCT-2000; 2000US-237172P.  
 PR 02-OCT-2000; 2000US-237173P.  
 PR 02-OCT-2000; 2000US-237278P.  
 PR 02-OCT-2000; 2000US-237294P.  
 PR 02-OCT-2000; 2000US-237295P.  
 PR 02-OCT-2000; 2000US-237316P.  
 PR 03-OCT-2000; 2000US-237425P.  
 PR 03-OCT-2000; 2000US-237598P.

PR 03-OCT-2000; 2000US-237604P.  
 PR 03-OCT-2000; 2000US-237606P.  
 PR 03-OCT-2000; 2000US-237608P.  
 PR 01-NOV-2000; 2000US-244867P.  
 PR 01-NOV-2000; 2000US-245084P.  
 XX (AVAL-) AVALON PHARM.  
 PI Young PE, Augustus M, Carter KC, Ehner R, Endress G, Horrihan S,  
 PI Soppet DR, Weaver Z;  
 DR WPI; 2002-188264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a  
 PT chemical agent to be tested for anti-neoplastic activity, and  
 PT determining a change in expression of a gene of a signature gene set -  
 claim 1; SEQ ID 5614; 44bp; English.

XX The present invention describes a method (M1) for screening for an  
 CC anti-neoplastic agent. The method involves exposing cells to a chemical  
 CC agent to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 CC to ABL70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytosarcoma  
 CC activity and can be used in gene therapy. M1 can be used for screening  
 CC an anti-neoplastic agent, and can be used for producing a product which  
 CC is the data collected with respect to the anti-neoplastic agent as a  
 CC result of M1, and the data is sufficient to convey the chemical  
 CC structure and/or properties of the agent. M1 can be used in the  
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer  
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
 CC carcinoma, papillary carcinoma and Wilms' tumour.

XX Sequence 532 BP; 110 A; 183 C; 113 G; 126 T; 0 other;

XX Query Match 35.0%; Score 508.4; DB 24; Length 532;

XX Best Local Similarity 99.4%; Pred. No. 6e-106;

XX Matches 531; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 712 AACCAATGATATGCTGAGACTGAGACCGGGCCAGAGACCTTGCAGAGTACTCTGA 771  
 DB 1 AACCAATGATATGCTGAGACTGAGACCGGGCCAGAGACCTTGCAGAGTACTCTGA 60  
 QY 772 GGGCCCCCTGCTTGTGAGAGAGACCTTCCAGGATCTCTCTGGGGGTATTACCCCTGT 831  
 DB 61 GGGCCCCCTGCTTGTGAGAGAGACCTTCCAGGATCTCTCTGGGGGTATTACCCCTGT 120  
 QY 832 GGGCTGCGCCAGATCCAGCTGTCTACACCGAGATCTGCAATTCATCTCTGATCAAT 891  
 DB 121 GGGCTGTG-CAGGATTCAGCTGTCTACACCGAGATCTGCAATTCATCTCTGATCAAT 179  
 QY 892 AAAGTCATAGCGCTCCAGTATCCAGATGCTACGCTCCAGCTGATCCAGATGTTATGCTC 951  
 DB 180 AAAGTCATAGCGCTCCAGTATCCAGATGCTACGCTCCAGTATCCAGATGTTATGCTC 239  
 QY 952 CTGCTGATCCAGATCCAGAGGCTCCATGCTTCATCTCTTCTCCCGAGTGGCTGAA 1011  
 DB 240 CTGCTGATCCAGATCCAGAGGCTCCATGCTTCATCTCTTCTCCCGAGTGGCTGAA 299  
 QY 1012 CTCTCCCTTGTCTGCACTGTTCAACCTCTGCGCGCCCTCCAGACCTTAACATCTGCC 1071  
 DB 300 CTCTCCCTTGTCTGCACTGTTCAACCTCTGCGCGCCCTCCAGACCTTAACATCTGCC 359  
 QY 1072 CTCTCACTCATTTCCCTCCCTATTCCTCTGCTGTACTGAAAGCTGAAATGCAGG 1131  
 DB 360 CTCTCACTCATTTCCCTCCCTATTCCTCTGCTGTACTGAAAGCTGAAATGCAGG 419  
 QY 1132 AAGTGTGGCAAGGTTTATTCGAGAGAGCAGAGCCGCTATACCCAGCTCTGA 1191

Db 420 AAGTGTGGCAAAAGTTTATTCAGAGAGCCAGGAACCGGTCATCATCCAGCTCTGA 479  
QY 1192 GAGAGATTAATGAGGAGTACCAACCTGATCTCTGCGCACTCCCGCTGTGG 1245  
Db 480 GAGAGATTAATGAGGAGTACCAACCTGATCTCTGCGCACTCCCGCTGTGG 532

## RESULT 10

ABL82556  
ID ABL82556 standard; cDNA; 532 BP.

AC ABL82556;

DT 17-MAY-2002 (first entry)

DE Human ovarian cancer related cDNA clone SEQ ID NO:5534.

XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

OS Homo sapiens.

PN WO200192581-A2.

PD 06-DEC-2001.

PF 29-MAY-2001; 2001MO-US17756.

PR 26-MAY-2000; 2000US-207484P.

XX (COR1-) CORIXA CORP.

PI Algate PA, Harlocker SL, Jones R;

XX WPI; 2002-122075/16.

XX Composition for therapy and diagnosis of ovarian cancer comprising  
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
PT polypeptide, antibody specific to polypeptide or T cell expressing  
PT polypeptide -

XX Claim 1; SEQ ID 5534; 489pp; English.

XX The present invention describes a composition (I) comprising: carriers  
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour  
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to  
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell  
CC population of (II), or antigen presenting cells that express (II).  
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to  
CC (S1) can be used for detecting ovarian cancer in a patient's biological  
CC sample preferably serum or ovarian tissue. The method comprises  
CC contacting a biological sample from a patient with (IV), detecting the  
CC amount of polynucleotide hybridising to (IV) and comparing the amount to  
CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
CC patient, where the amount of polynucleotide hybridising to (IV) is  
CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
CC (III) and/or (II) is useful for stimulating and/or expanding T cells  
CC specific for an ovarian tumour protein comprising contacting T cells  
CC with (III) or (II). (III) is useful in design and preparation of  
CC ribozyme molecules for inhibiting expression of the tumour polypeptides  
CC and proteins in tumour cells; and to isolate a full length gene from a  
CC suitable library e.g., a tumour cDNA library using well known  
CC techniques.

XX Sequence 532 BP; 110 A; 103 C; 113 G; 126 T; 0 other;

XX Query Match 35.0%; Score 508.4; DB 24; Length 532;  
XX Best Local Similarity 99.4%; Pred. No. 6e-106;  
XX Matches 531; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 712 AACACATGATATGCTGACCTGACCGGCGCCAGGACCTTCCAGAGTACTTCGGA 771  
Db 1 AACACATGATATGCTGACCTGACCGGCGCCAGGACCTTCCAGAGTACTTCGGA 60

QY 772 GGGCCCTGATGCTGATGAGAGACCCCTCCAAAGGATCTCTCGGAGGATTTTACCCCTGT 831  
Db 61 GGGCCCTGATGCTGATGAGAGACCCCTCCAAAGGATCTCTCGGAGGATTTTACCCCTGT 120  
QY 832 GGGCTGCGCCAGATCCAGCTGCTCTACACCCAGATCTGCAAAATACATGCTTGATCAAT 891  
Db 121 GGGCTGCG -CGAGATCCAGCTGCTCTACACCCAGATCTGCAAAATACATGCTTGATCAAT 179  
QY 892 AAAATCATAGCTCCAACTGATCCAGATCTGATCCAGCTGATCCAGATCTGATCTGCT 951  
Db 180 AAAATCATAGCTCCAACTGATCCAGATCTGATCCAGCTGATCCAGATCTGATCTGCT 239  
QY 952 CTGCTATTCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1011  
Db 240 CTGCTATTCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 299  
QY 1012 CTCTCCCTGATGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1071  
Db 300 CTCTCCCTGATGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 359  
QY 1072 CTCTCACTTCATTCCTCCCACTTCCTCCCACTTCCTCCCACTTCCTCCCACTTCCT 1131  
Db 360 CTCTCACTTCATTCCTCCCACTTCCTCCCACTTCCTCCCACTTCCTCCCACTTCCT 419  
QY 1132 AAGTGTGGCAAAAGTTTATTCAGAGAGCCAGGAACCGGTCATCATCCAGCTCTGA 1191  
Db 420 AAGTGTGGCAAAAGTTTATTCAGAGAGCCAGGAACCGGTCATCATCCAGCTCTGA 479  
QY 1192 GAGAGATTAATGAGGAGTACCAACCTGATCTCTGCGCACTCCCGCTGTGG 1245  
Db 480 GAGAGATTAATGAGGAGTACCAACCTGATCTCTGCGCACTCCCGCTGTGG 532

## RESULT 11

ABL82657/c  
ID ABL82657 standard; cDNA; 454 BP.

AC ABL82657;

DT 17-MAY-2002 (first entry)

DE Human ovarian cancer related cDNA clone SEQ ID NO:5635.

XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

OS Homo sapiens.

PN WO200192581-A2.

PD 06-DEC-2001.

PF 29-MAY-2001; 2001MO-US17756.

PR 26-MAY-2000; 2000US-207484P.

XX (COR1-) CORIXA CORP.

PI Algate PA, Harlocker SL, Jones R;

XX WPI; 2002-122075/16.

XX Composition for therapy and diagnosis of ovarian cancer comprising  
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
PT polypeptide, antibody specific to polypeptide or T cell expressing  
PT polypeptide -

XX Claim 1; SEQ ID 5635; 489pp; English.

XX The present invention describes a composition (I) comprising: carriers  
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour  
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to







DB 361 TGTAAGCTGAAATGACAGAGAGTGTCGCAAAAGTTTATTCACAGAGAGCAGAG 420  
QY 1170 CCGGTCATCAACCCAGCCTCTGAGAGCACT 1198  
DB 421 CCGGTCATCAACCCAGCCTCTGAGAGCACT 449

## RESULT 13

ABL79800/c  
ID ABL79800 standard; cDNA; 470 BP.

AC ABL79800;

DT 17-MAY-2002 (first entry)

DE Human ovarian cancer related cDNA clone SEQ ID NO:2778.

KM Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

OS Homo sapiens.

PN WO200192581-A2.

PD 06-DEC-2001.

PF 29-MAY-2001; 2001WO-US17756.

PR 26-MAY-2000; 2000US-207484P.

PA (CORI-) CORIXA CORP.

PI Algate PA, Harlocker SL, Jones R;

DR WPI; 2002-122075/16.

PT Composition for therapy and diagnosis of ovarian cancer comprising  
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
PT polypeptide, antibody specific to polypeptide or T cell expressing  
PT polypeptide

Claim 1; SEQ ID 2778; 489pp; English.

The present invention describes a composition (I) comprising: carriers  
CC and immunostimulants; and a polypeptide (II) of a ovarian tumor  
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to  
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell  
CC population of (II), or antigen presenting cells that express (II).  
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridizes to  
CC (S1) can be used for detecting ovarian cancer in a patient's biological  
CC sample preferably serum or ovarian tissue. The method comprises  
CC contacting a biological sample from a patient with (IV), detecting the  
CC amount of polynucleotide hybridizing to (IV) and comparing the amount to  
CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
CC patient, where the amount of polynucleotide hybridizing to (IV) is  
CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
CC (III) and/or (II) is useful for stimulating and/or expanding T cells  
CC specific for an ovarian tumor protein comprising contacting T cells  
CC with (III) or (II). (III) is useful in design and preparation of  
CC ribosome molecules for inhibiting expression of the tumor polypeptides  
CC and proteins in tumor cells; and to isolate a full length gene from a  
CC suitable library e.g., a tumor cDNA library using well known  
CC techniques.

CC Sequence 470 BP; 127 A; 96 C; 142 G; 105 T; 0 other;

Query Match 30.6%; Score 445.4; DB 24; Length 470;

Best Local Similarity 99.4%; Pred. No. 1.2e-91;

Matches 468; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 974 GCTCATGCTCACTCTTCTCTCCAGTCGCTGAACTCTCCCTGCTGCACTGTT 1033  
DB 470 GCTCATGCTCACTCTTCTCTCCAGTCGCTGAACTCTCCCTGCTGCACTGTT 411

QY 1034 CAAACCTTCGCGGCGCTCCACACCTCTAAACATCTCCCTCTACCTCATTTCCCCACCT 1093  
DB 410 CAAACCTTCGCGGCGCTCCACACCTCTAAACATCTCCCTCTACCTCATTTCCCCACCT 351  
QY 1094 ATCCCATTTCTCTGCTCTGACTGAACTGAAATGCAAGAAAGTGTGGCAAGGTTATTC 1153  
DB 350 ATCCCATTTCTCTGCTCTGACTGAACTGAAATGCAAGAAAGTGTGGCAAGGTTATTC 291  
QY 1154 CAGAGAAAGCCAGAGAACCGGCTATCAACCCAGCTCTGAGAGAGTAACTGGGGTCAACCA 1213  
DB 290 CAGAGAAAGCCAGAGAACCGGCTATCAACCCAGCTCTGAGAGAGTAACTGGGGTCAACCA 231  
QY 1214 ACCTGACTTCCTGCACTCCCGGCTGTGTGCTTTGGGCAACCAAGTCCCTCTCTG 1273  
DB 230 ACCTGACTTCCTGCACTCCCGGCTGTGTGCTTTGGGCAACCAAGTCCCTCTCTG 172  
QY 1274 AACCTGACTTCCTGCACTCCCGGCTGTGTGCTTTGGGCAACCAAGTCCCTCTGAGACATGTT 1333  
DB 171 AACCTGACTTCCTGCACTCCCGGCTGTGTGCTTTGGGCAACCAAGTCCCTCTGAGACATGTT 112  
QY 1334 GTGAGAGACATGATATTAACATGTGTATGTAATCTTCAT-GTGAATGTCATGTAAGGC 1392  
DB 111 GTGAGAGACATGATATTAACATGTGTATGTAATCTTCATGATGTCATGTAAGGC 52  
QY 1393 TTACACAGTGGGTGGTGAGTTCTGACTAAAGTTACCTGTTCTGCTGAAA 1443  
DB 51 TTACACAGTGGGTGGTGAGTTCTGACTAAAGTTACCTGTTCTGCTGAAA 1

## RESULT 14

ABL79803/c  
ID ABL79803 standard; cDNA; 468 BP.

AC ABL79803;

DT 17-MAY-2002 (first entry)

DE Human ovarian cancer related cDNA clone SEQ ID NO:2781.

KM Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

OS Homo sapiens.

PN WO200192581-A2.

PD 06-DEC-2001.

PF 29-MAY-2001; 2001WO-US17756.

PR 26-MAY-2000; 2000US-207484P.

PA (CORI-) CORIXA CORP.

PI Algate PA, Harlocker SL, Jones R;

DR WPI; 2002-122075/16.

PT Composition for therapy and diagnosis of ovarian cancer comprising  
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
PT polypeptide, antibody specific to polypeptide or T cell expressing  
PT polypeptide

Claim 1; SEQ ID 2781; 489pp; English.

The present invention describes a composition (I) comprising: carriers  
CC and immunostimulants; and a polypeptide (II) of a ovarian tumor  
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to  
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell  
CC population of (II), or antigen presenting cells that express (II).  
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridizes to  
CC (S1) can be used for detecting ovarian cancer in a patient's biological

CC sample preferably serum or ovarian tissue. The method comprises  
CC connecting a biological sample from a patient with (IV), detecting the  
CC amount of polynucleotide hybridising to (IV) and comparing the amount to  
CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
CC patient, where the amount of polynucleotide hybridising to (IV) is  
CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
CC (III) and/or (II) is useful for stimulating and/or expanding T cells  
CC specific for an ovarian tumour protein comprising contacting T cells  
CC with (III) or (I). (II) is useful in design and preparation of  
CC ribozyme molecules for inhibiting expression of the tumour polypeptides  
CC and proteins in tumour cells; and to isolate a full length gene from a  
CC suitable library e.g., a tumour cDNA library using well known  
CC techniques.

Sequence 468 BP; 118 A; 92 C; 152 G; 106 T; 0 other;

Query Match	30.2%;	Score 438.4;	DB 24;	Length 468;
Best Local Similarity	99.6%;	Pred. No. 4.7e-90;		
Matches 450; Conservative	0;	Mismatches 1;	Indels 1;	Gaps 1;

QY	922	TAGAGCTCAGAGCGATACAAATGTTATGCTCCGGGTATGCCAGATATGCCAGAGCTCATC	981
Db	452	TAGGCTTCAGCTGATGCAGATGTATATGCTCTGCTGATCCAGATGCCAGAGCTCATC	393
QY	982	GTCCATCTCTTCTCCCGAGTGGCTGAACCTCTCCCTGTGTGACACTGTTCAAACCTC	1044
Db	392	GTCCATCTCTTCTCTCCCGAGTGGCTGAACCTCTCCCTGTGTGACACTGTTCAAACCTC	333
QY	1042	TGCGCGCTCCAACTCTAAACATCTCCCTCACTCATCTATTCGCCCACTATGCCCAT	1107
Db	332	TGCGCGCTCCAACTCTAAACATCTCCCTCACTCATCTATTCGCCCACTATGCCCAT	273
QY	1102	TCTCTGCTCTACTGAGAGCTGAATATGACAGAAAGTGTGTGCAAGATTTATTCAGAGAG	1161
Db	272	TCTCTGCTCTACTGAGAGCTGAATATGACAGAAAGTGTGTGCAAGATTTATTCAGAGAG	213
QY	1162	CCAGAGAACCGGTATCATCACTCCAGACCTCTGAGAGCAGTTACTGGGGTATCCCAACTGACT	1222
Db	212	CCAGAGAACCGGTATCATCACTCCAGACCTCTGAGAGCAGTTACTGGGGTATCCCAACTGACT	153
QY	1222	TCTCTGCTCACTCCCGCTGTGTGACTTTGTGGGCAAGCCAAATGTGCTCTCTGAACCTCAG	1287
Db	152	TCTCTGCTCACTCCCGCTGTGTGACTTT-GGCAAGCCAAATGTGCTCTCTGAACCTCAG	94
QY	1282	TTTCTCTCATCTGCAAAATGGGAAACAATGACGTGCTTACTCTTGAGACATGTGTGAGAG	1344
Db	93	TTTCTCTCATCTGCAAAATGGGAAACAATGACGTGCTTACTCTTGAGACATGTGTGAGAG	34
QY	1342	ACTATGATATTAACATGTGTATGAAATCTTCA	1373
Db	33	ACTATGATATTAACATGTGTATGAAATCTTCA	2

RESULT 15	
ABL82197/G	
ID	ABL82197 standard; cDNA; 399 BP.
XX	
AC	ABL82197;
XX	
DT	17-MAY-2002 (first entry)
XX	
DE	Human ovarian cancer related cDNA clone SEQ ID NO:5175.
XX	
KM	Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss
XX	
OS	Homo sapiens.
XX	
PN	WO200132581-A2.
XX	
PD	06-DEC-2001.
XX	
PF	29-MAY-2001; 2001WO-US17756.
XX	

PR 26-MAY-2000; 2000US-207484P  
XX  
XX (CORI-) CORIXA CORP.  
PA  
PI Algate PA, Hatlockker SL,  
XX  
XX  
XX  
XX  
XX WPI; 2002-122075/16.  
DR

PT Composition for therapy and diagnosis of ovarian cancer comprising  
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
PT polypeptide, antibody specific to polypeptide or T cell expressing  
PT polypeptide -  
XX  
PS Claim 1, SEQ ID 5175; 489pp; English.

The present invention describes composition (i) comprising: carriers and immunostimulants; and a polypeptide (ii) of an ovarian tumour polypeptide encoded by a polynucleotide (iii) having a cDNA sequence (Si). From the 10912 nucleotide sequences as given in ABL77023 to ABL87934, (iii) encoding (ii) having a sequence (S2), a T cell population of (ii), or antigen presenting cells that express (ii). (i) has cytostatic activity. An oligonucleotide (iv) that hybridises to (Si) can be used for detecting ovarian cancer in a patient's biological sample preferably serum or ovarian tissue. The method comprises contacting a biological sample from a patient with (iv), detecting the amount of polynucleotide hybridising to (iv) and comparing the amount to a predetermined cutoff value and thereby detecting ovarian cancer in the patient, where the amount of polynucleotide hybridising to (iv) is detected preferably by polymerase chain reaction (PCR). (i) comprising (iii) and/or (ii) is useful for stimulating and/or expanding T cells specific for an ovarian tumour protein comprising and/or contacting T cells with (iii) or (ii). (iii) is useful in design and preparation of ribozyme molecules for inhibiting expression of the tumour polypeptides and proteins in tumour cells; and to isolate a full length gene from a suitable library e.g., a tumour cDNA library using well known techniques.

5Q Sequence 399 BP; 106 A; 86 C; 110 G; 97 T; 0 other;

Query Match	26.4%	Score 383.8	DB 24	length 399
Best Local Similarity	99.2%	Pred. No. 1.1e-77		
Matches 356; Conservative	0	Mismatches 2	Indels 1	Gaps 1

Qy	1046	GGCTCCACACTCTGAAACATCTCCCTCCACCTGAACTGCTGACCTGACCTGCTG	1108
Db	359	GGCTTCACACTCTTAAACATCTCCCTCTACCTCATCTCCCACTATCCCATCTTC	340
Qy	1106	TGCTGTACTGAAGCTGAAATGACAGSAGTGTGGCAAGGTTTATTCACAGAGCCAG	1165
Db	339	TGCTGTACTGAAGCTGAAATGACAGSAGTGTGGCAAGGTTTATTCACAGAGCCAG	280
Qy	1166	GAAAGCGGTCAATCAACCGCTCTGAGAGAGTTACTGGGGTCAACCCAACTGACTTCT	1225
Db	279	GAAAGCGGTCAATCAACCGCTCTGAGAGAGTTACTGGGGTCAACCCAACTGACTTCT	220
Qy	1226	CTGCACATCCCCGCTGTGTGACTTTGGGCAAGCCAAAGGCCCTCTGTAACTCAGTTTC	1285
Db	219	CTGCACATCCCCGCTGTGTGACTTTGGGCAAGCCAAAGGCCCTCTGTAACTCAGTTTC	160
Qy	1286	CTCATCTGCAAAATGGAGACATGACGTGCTACCTCTTAACACATGTTGTGGAGAGACTA	1345
Db	159	CTCATCTGCAAAATGGAGACATGACGTGCTACCTCTTAACACATGTTGTGGAGAGACTA	100
Qy	1346	TGATTTAACATGTGTATGTAATCTTCAT-GTATTTGATGTAAGGCTTAAACACAGTGG	1405
Db	99	TGATTTAACATGTGTATGTAATCTTCATGATGTGATGTGACATGTAAGGCTTAAACACAGTGG	40
Qy	1405	GTGTGAGTTCGACTAAGGTAACCTTTGTGCGTAAA	1443
Db	39	GTGTGAGTTCGACTAAGGTAACCTTTGTGCGTAAA	1

Wed Nov 26 18:23:55 2003

us-10-021-368-2.rng

Page 15

Job time : 430 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: November 25, 2003, 04:52:28 ; Search time 5442 Seconds

(without alignments)  
10930.282 Million cell updates/sec

Title: US-10-021-368-2

Perfect score: 1454  
Sequence: 1 ACCGAGCGGACAGACACAGGC.....GTCGTGAAAAAAAAAAAA 1454

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: GenBank:\*  
2: gb\_ba:\*  
3: gb\_hvg:\*  
4: gb\_in:\*  
5: gb\_om:\*  
6: gb\_ov:\*  
7: gb\_pat:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_xo:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pi:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hvg\_hum:\*  
31: em\_hvg\_inv:\*  
32: em\_hvg\_other:\*  
33: em\_hvg\_mus:\*  
34: em\_hvg\_pin:\*  
35: em\_hvg\_rnd:\*  
36: em\_hvg\_mam:\*  
37: em\_hvg\_vrt:\*  
38: em\_sy:\*  
39: em\_hvg\_hum:\*  
40: em\_hvg\_mus:\*  
41: em\_hvg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1454	100.0	1454	6	AR000484
2	1454	100.0	1454	6	AR062140
3	1454	100.0	1454	6	AR119561
4	1454	100.0	1454	6	AF024605
5	1301.6	89.5	1542	9	BC002710
6	829.4	57.0	851	6	AX360072
7	686	47.2	5574	9	AF055481
8	673.4	46.3	13233	9	AC011473
9	663.4	46.3	230000	9	AF243527
10	663.8	45.7	200792	2	AC130782
11	662.2	45.5	178504	2	AC140096
12	633.6	43.6	142394	2	AC140096
13	608.4	35.0	532	6	AX331169
14	508.4	35.0	532	6	AX331572
15	508.4	35.0	532	6	AX335105
16	500	34.4	1380	10	BC002100
17	483.8	33.3	176647	2	AC130188
18	376	25.3	396	6	AX311778
19	376	25.3	396	6	AX332667
20	376	25.9	396	6	AX333980
21	376	25.9	396	6	AX333980
22	376	25.9	396	6	AX333980
23	376	25.9	396	6	AX333980
24	376	25.9	396	6	AX333980
25	376	25.9	396	6	AX333980
26	267.2	18.4	283	6	AX334883
27	203.4	14.0	205	6	AX335313
28	203.4	14.0	205	6	AX335313
29	173.8	12.0	1091	6	AX208857
30	165.8	11.4	1166	6	AX198293
31	164	11.3	1106	6	AX152173
32	164	11.3	1106	6	AX152173
33	164	11.3	1106	6	AX152173
34	164	11.3	1106	6	AX152173
35	164	11.3	1106	6	AX152173
36	164	11.3	1106	6	AX152173
37	164	11.3	1106	6	AX152173
38	164	11.3	1106	6	AX152173
39	164	11.3	1106	6	AX152173
40	164	11.3	1106	6	AX152173
41	164	11.3	1106	6	AX152173
42	164	11.3	1106	6	AX152173
43	164	11.3	1106	6	AX152173
44	164	11.3	1106	6	AX152173
45	164	11.3	1106	6	AX152173

#### ALIGNMENTS

RESULT 1  
LOCUS AR000484  
DEFINITION Sequence 2 from patent US 5736377.  
ACCESSION AR000484  
VERSION AR000484.1 GI:3963015  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1454)  
AUTHORS Band, V.  
TITLE NES-1 polypeptides, DNA, and related molecules and methods  
JOURNAL Patent: US 5736377-A 2 07-Apr-1998;  
FEATURES Location/Qualifiers



QY 181 GGGCTGCTCCCCCAAAAGACACGCGCTTGGACCCCGAAGCTATGGCCCGCGGCG 240  
 DB 181 GGGCTGCTCCCCCAAAAGACACGCGCTTGGACCCCGAAGCTATGGCCCGCGGCG 240  
 QY 241 CGCGGCTGCGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCT 300  
 DB 241 CGCGGCTGCGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCT 300  
 QY 301 GTCCGCTGCGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCT 360  
 DB 301 GTCCGCTGCGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCT 360  
 QY 361 GTCCGCTGCGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCT 420  
 DB 361 GTCCGCTGCGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCT 420  
 QY 421 GCTCGAGTAGGAGATGATCACTGCTGCTTCTTCAAGGAGAGAGAGAGAGAGAGAG 480  
 DB 421 GCTCGAGTAGGAGATGATCACTGCTGCTTCTTCAAGGAGAGAGAGAGAGAGAGAG 480  
 QY 481 GATGAGACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 DB 481 GATGAGACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 QY 541 CGGCGCTGCGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCT 600  
 DB 541 CGGCGCTGCGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCTTGGACGCT 600  
 QY 601 TGGGCGACACG 660  
 DB 601 TGGGCGACACG 660  
 QY 661 ACTATCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 DB 661 ACTATCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 QY 721 ATATGCTGCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 DB 721 ATATGCTGCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 QY 781 GTCTGCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
 DB 781 GTCTGCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
 QY 841 CAGATCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 DB 841 CAGATCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 QY 901 CGCTGCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 DB 901 CGCTGCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 QY 961 CAGATCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 DB 961 CAGATCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 QY 1021 TGTCTGCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 DB 1021 TGTCTGCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 QY 1081 CATTCTGCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
 DB 1081 CATTCTGCTGAGACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
 QY 1141 CAAAGGTTTATTCAG 1200  
 DB 1141 CAAAGGTTTATTCAG 1200  
 QY 1201 CTGGGAGTCAACCACTGATCTTCTCTGCGACCTCCCGCTGCTGCTTGGAGAGAGAG 1260  
 DB 1201 CTGGGAGTCAACCACTGATCTTCTCTGCGACCTCCCGCTGCTGCTTGGAGAGAGAG 1260

QY 1261 AGTGCCTCTCTGAGACCTGAGATTTCCTCATCTGCAAAATAGGAAAGAGAGAGAG 1320  
 DB 1261 AGTGCCTCTCTGAGACCTGAGATTTCCTCATCTGCAAAATAGGAAAGAGAGAGAG 1320  
 QY 1321 TCTTGAACATGTTGTGAG 1380  
 DB 1321 TCTTGAACATGTTGTGAG 1380  
 QY 1381 GTCATGTGAAGCTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 DB 1381 GTCATGTGAAGCTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 QY 1441 AAAAAAAAAAAAAA 1454  
 DB 1441 AAAAAAAAAAAAAA 1454  
 RESULT 3  
 AR119561  
 LOCUS AR119561 1454 bp DNA linear PAT 16-MAY-2001  
 DEFINITION Sequence 2 from patent US 6153387.  
 ACCESSION AR119561  
 VERSION AR119561.1 GI:14102260  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 1454)  
 AUTHORS Band, V.  
 TITLE NBS-1 polypeptides, DNA, and related molecules and methods  
 JOURNAL Patent: US 6153387-A 2 28-NOV-2000;  
 FEATURES  
 source 1..1454  
 BASE COUNT 289 a 481 c 377 g 307 t  
 ORIGIN  
 Query Match 100.0%; Score 1454; DB 6; Length 1454;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-292;  
 Matches 1454; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ACCGAGCGGACAGACACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
 DB 1 ACCGAGCGGACAGACACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
 QY 61 GCGAGCTCCAGATCTCTGCGCATGAGAGCTCCGACCTCCACTCTTCGCGCCTTGGC 120  
 DB 61 GCGAGCTCCAGATCTCTGCGCATGAGAGCTCCGACCTCCACTCTTCGCGCCTTGGC 120  
 QY 121 GCCCGGCTCTGCGGAG 180  
 DB 121 GCCCGGCTCTGCGGAG 180  
 QY 181 GCGCTGCTCCCAAAACAGACAGCGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
 DB 181 GCGCTGCTCCCAAAACAGACAGCGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
 QY 241 CGGCGCTGCGAGCGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
 DB 241 CGGCGCTGCGAGCGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
 QY 301 GTCTGGTGAACAG 360  
 DB 301 GTCTGGTGAACAG 360  
 QY 361 GCTCGAGTAGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
 DB 361 GCTCGAGTAGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
 QY 421 CGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
 DB 421 CGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480



```

Db      301  GTCTGTGTCACAGAGTTGGTGTGACGGCCGCGACCTGCGGAAACAGGCCACTGTGG 360
Qy      361  GCTCGAGTAGGGATGATCACTGTGCTTCTTCAAGGGCAGAGCACTCCGCGAGACT 420
Db      361  GCTCGAGTAGGGATGATCACTGTGCTTCTTCAAGGGCAGAGCACTCCGCGAGACT 420
Qy      421  CGCTCTGTGTCCTCCCAAGTACACACAGGGCTCAGGGCCCTCTCTCCCAAGGGCAG 480
Db      421  CGCTCTGTGTCCTCCCAAGTACACACAGGGCTCAGGGCCCTCTCTCCCAAGGGCAG 480
Qy      481  GATAGACGATCTCATGTGTCTAAAGCTGGCCAGGGCCGCTAGTGGCCGGGGCCGCT 540
Db      481  GATAGACGATCTCATGTGTCTAAAGCTGGCCAGGGCCGCTAGTGGCCGGGGCCGCT 540
Qy      541  CGGGCCCTGTGAGTTCTTACCTGCTGCTCAGCCCGGAGACCAAGTCCAGATTGTGCT 600
Db      541  CGGGCCCTGTGAGTTCTTACCTGCTGCTCAGCCCGGAGACCAAGTCCAGATTGTGCT 600
Qy      601  TGGGGGACACAGCCGCGCGCGGAGAGTACACACAGGGCCCTGACCTGCTCCAGAC 660
Db      601  TGGGGGACACAGCCGCGCGCGGAGAGTACACACAGGGCCCTGACCTGCTCCAGAC 660
Qy      661  ACTATCTGTAGCCCTTAAAGAGTGTAGGTCTTCTTACCTGCGCTGCTCAACACAT 720
Db      661  ACTATCTGTAGCCCTTAAAGAGTGTAGGTCTTCTTACCTGCGCTGCTCAACACAT 720
Qy      721  ATATGCTGTGACCTGGACCGCGCGGAGAGACCTTTCACAGATCTTGTGAGGGCCCT 780
Db      721  ATATGCTGTGACCTGGACCGCGCGGAGAGACCTTTCACAGATCTTGTGAGGGCCCT 780
Qy      781  GTCTGTGACGAGACCTTCCAGAGGATCTCTGCTGGGGGTGTAAACCTGTGGCTTGC 840
Db      781  GTCTGTGACGAGACCTTCCAGAGGATCTCTGCTGGGGGTGTAAACCTGTGGCTTGC 840
Qy      841  CAGATATCCAGCTGTCTACACCCAGATCTGCAATATGCTGATCTGATTAAGTCATA 900
Db      841  CAGATATCCAGCTGTCTACACCCAGATCTGCAATATGCTGATCTGATTAAGTCATA 900
Qy      901  CGCTCAACTGATCTCAGATCTCAGCTCAGCTGATCTCAGATCTTGTGCTGCTGAT 960
Db      901  CGCTCAACTGATCTCAGATCTCAGCTCAGCTGATCTCAGATCTTGTGCTGCTGAT 960
Qy      961  CAGATGCTCAGAGGCTTCATCTGCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Db      961  CAGATGCTCAGAGGCTTCATCTGCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Qy      1021  TGTCTGCACTGTTCAGAACTCTGCGCGCCCTCCACACTTAAACATCTCCCTCTCA 1080
Db      1021  TGTCTGCACTGTTCAGAACTCTGCGCGCCCTCCACACTTAAACATCTCCCTCTCA 1080
Qy      1081  CATTCCTCCCACTTATCCCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db      1081  CATTCCTCCCACTTATCCCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Qy      1141  CAAAGGTTTATTCAGAGAGAGCAGAGAGCCGCTCATACACAGCTCTGAGAGCAG 1200
Db      1141  CAAAGGTTTATTCAGAGAGAGCAGAGAGCCGCTCATACACAGCTCTGAGAGCAG 1200
Qy      1201  CTGGGGTCACTCACTGACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db      1201  CTGGGGTCACTCACTGACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Qy      1261  AGTCCCTCTCTGAACTCTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1320
Db      1261  AGTCCCTCTCTGAACTCTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1320
Qy      1321  TCTTAAAGCATGTTGTGAGAGAGCAGTAAATTAACATGTTGATTAATCTTCA 1380
Db      1321  TCTTAAAGCATGTTGTGAGAGAGCAGTAAATTAACATGTTGATTAATCTTCA 1380
Qy      1381  GTATGTAAAGGCTTAAACAGAGTGGGTGTGAGTTTGTGACCTAAAGTTACTGT 1440
Db      1381  GTATGTAAAGGCTTAAACAGAGTGGGTGTGAGTTTGTGACCTAAAGTTACTGT 1440

```

Qy 1441 AAAAAAAAAAAAAA 1454  
Db 1441 AAAAAAAAAAAAAA 1454

# RESULT 5 LOCUS DEFINITION

BC002710 1542 bp mRNA linear PRI 12-JUL-2001  
Homo sapiens, kallikrein 10, clone MGC:3667 IMAGE:3632557, mRNA,  
complete cds.

# ACCESSION VERSION KEYWORDS SOURCE ORGANISM

# REFERENCE AUTHORS TITLE JOURNAL

1 (bases 1 to 1542)  
Strausberg, R.  
Direct Submission  
Submitted (05-FEB-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Shevchenko, Y., Welteby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakeley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,  
McConnell, J., Pearson, R., Snyder, B., Stancitop, S., Thomas, P.J.,  
Tongson, E.B., Touchman, D.W., Yauzeon, C., Vogt, J.L., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 12 Row: a Column: 24  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 4506156.  
Location/Qualifiers

# FEATURES source

1..1542  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="locusid:5655"  
/db\_xref="taxon:9606"  
/clone="MGC:3667 IMAGE:3632557"  
/cfeature="UTR", endometrium adenocarcinoma"  
/lab\_host="DH10B-R"  
/note="Vector: pOT87"  
229..1059  
/codon\_start=1  
/product="kallikrein 10"  
/protein\_id="AA02710.1"  
/db\_xref="gi:12803745"  
/translation="NSRPHILSLPASGALAKLFLMAQMAEALPONDYRLD  
PEAVGAPCARQSPQVSLFNGLSFGAGVLVDQSVLPAHGNLPAARVGDHLL  
LLOGEQLRTTRSVVHPRKYQSGSLIPRTBEDHMLAKLAPVPGRRVALQIPY  
RCAPGQOCOVAGGTTAARVYKNGLCSSSTTISPRECEVFEGVATNNITCGL  
DRGDPQSDSGSLVCDSTLQIISWGYPPCSAQRAPVYVYICRYMWINKIVASN

# CDS

BASE COUNT 317 a 517 c 404 g 304 t  
ORIGIN







Df		5029	TCCCAAGTGGGCTGAACCTCCTCCCCTGTGTGCATGTTCAAACCGTCGCGGCCCTCCACA	5086
Oy		1056	CCTCTAAACACTTCCCCCTCTCACTCATATTCGCCCACTATCCCAATTCTTGCTGTACT	1115
Df		5089	CCTCTAAACACTTCCCCCTCTCACTCATATTCGCCCACTATCCCAATTCTTGCTGTACT	5148
Oy		1116	GAACTGAAAATCAGAGAATGAGTGGCGAAGGTTTATTCGAGGAAGCAAGAACCGGAGT	1175
Df		5149	GAACTGAAAATCAGAGAATGAGTGGCGAAGGTTTATTCGAGGAAGCAAGAACCGGAGT	5208
Oy		1176	ATCACCCAGCCTCTGAGAGCAGATTACTG399GTCAACCACCTGACTTCTCTGCACATCC	1235
Df		5209	ATCACCCAGCCTCTGAGAGCAGATTACTG399GTCAACCACCTGACTTCTCTGCACATCC	5266
Oy		1236	CCGCTGTGTGACTTTGGGGCAAGCCAGTGGCCCTCTCTBAACTCAGTTTCTTCATCTGCA	1295
Df		5269	CCGCTGTGTGACTTTGGGGCAAGCCAGTGGCCCTCTCTBAACTCAGTTTCTTCATCTGCA	5328
Oy		1296	AAATGGGAACAATGACGTCGCTCACTCTTAAACATGTTGTGAGAGACATGATATTAACA	1355
Df		5329	AAATGGGAACAATGACGTCGCTCACTCTTAAACATGTTGTGAGAGACATGATATTAACA	5388
Oy		1356	TGTGATCTTAAATCTTCATGTGATGTCTATGTAAGCTTAACACAGTGGGTGAGATTC	1415
Df		5389	TGTGATCTTAAATCTTCATGTGATGTCTATGTAAGCTTAACACAGTGGGTGAGATTC	5448
Oy		1416	TGACTAAAGATTACCTGTTGTCTGCA	1441
Df		5449	TGACTAAAGATTACCTGTTGTCTGCA	5474
RESULT 8				
AC011473/c		AC011473	132323 bp DNA linear PRI 12-OCT-2000	
LOCUS			Homo sapiens chromosome 19, BAC BC349142 (CTC-518B2), complete	
DEFINITION			sequence.	
ACCESSION		AC011473		
VERSION		AC011473.4	GI:10799392	
KEYWORDS		HTG.		
SOURCE				
ORGANISM			Homo sapiens (human)	
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS			Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
			1 (bases 1 to 132323)	
			Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Viswanathan,V.,	
			Burkhardt-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stillwagen,S.,	
			Pahan,H., Velasco,N., Do,L., Regala,W., Terry,A., Brower,A.,	
			Garnes,J., Danganan,L., Exler,A., Christensen,M., Georgescu,A.,	
			Avila,J., Liu,S., Andreise,T., Tranheim,M., Altix,C.,	
			Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R.,	
			Thomas,P., Quan,G., Krommiller,B., Ariellano,A., Sanders,C., Ow,D.,	
			Nolan,M., Truong,S., Kobayashi,A., Olsen,A.S. and Carraro,A.V.	
			Sequence analysis of chromosome 19q13.4	
			Unpublished	
TITLE			2 (bases 1 to 132323)	
JOURNAL			DOE Joint Genome Institute.	
REFERENCE			Direct Submission	
AUTHORS			Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint	
JOURNAL			Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
			3 (bases 1 to 132323)	
			Lamerdin,J.E.	
			Direct Submission	
			Submitted (12-OCT-2000) DOE Joint Genome Institute, Lawrence	
			Livermore National Laboratory, Livermore, CA 94550	
			On Oct 12, 2000 this sequence version replaced gi:7458725.	
			Map and sequence oriented from centromere to q-telomere. BC349142	
			(CTC-518B2) overlaps BAC BC85745 (CTB-147C22, AC011483) on the left	
			from bases 1 to 2,226 of this accession, and overlaps BAC BC892989	
			(CTD-1187F8, AC063977) on the right from bases 120,322 to 132,323	
			of this accession. Additional chromosome 19 map and sequence	
			information are available at:	
			http://www.bio.linn.gov/bio/gv/genome/genome.html.	
			Location/Qualifiers	
FEATURES				

source	1. 13223
gene	<p>/organism="Homo sapiens"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:9606"</p> <p>/chromosome="19"</p> <p>/map="19q13.4"</p> <p>/clone="CTC-519B2 (LNL clone name BC349142)"</p> <p>/clone_11b="CIT-BSPC"</p> <p>/complement(&lt;77. 1783)</p> <p>/gene="KLK8"</p> <p>/note="kallikrein 8 (neuropsin/ovasin); synonyms: TADG14, PSSS19, HNP, protease, serine, 19 (neuropsin/ovasin)"</p> <p>/complement(join(&lt;77. 339,505. 664,1179. 1256,1613). 1783)</p> <p>/gene="KLK8"</p> <p>/product="KLK8 mRNA"</p> <p>/complement(join(&lt;77. 339,505. 664,1179. 1248))</p> <p>/gene="KLK8"</p> <p>/note="potential molecular marker for ovarian carcinomas"</p> <p>/codon_start=1</p> <p>/product="kallikrein 8 (neuropsin/ovasin) (amino acids 1-164)"</p> <p>/protein_id="AAG2254.1"</p> <p>/db_xref="GI:10799393"</p> <p>/translation="MGRRPRPAATKTFMLLIGAWAGHSPAQEDKVLGHECOPHSQ PMQALFGQQLCGVYVGNWVLTAAHCKPKYVTLGHSYLNKDGPEQSLPVVQ SLPHCNSSDVEDHNDMLMQLRDSALGSKVKPISLADHCTQPGKCTVSGMGV TSPR"</p>
CDS	<p>382. 473</p> <p>/rpt_family="MSR1"</p> <p>/complement(1374. 1554)</p> <p>/rpt_family="MSR1"</p> <p>1723. 1754</p> <p>/rpt_family="(TCC)n"</p> <p>/complement(1791. 2238)</p> <p>/rpt_family="MSR1"</p> <p>/complement(2594. 9715)</p> <p>/gene="KLK9"</p> <p>/note="kallikrein 9; synonyms: KLK-L3, kallikrein-like protein 3"</p> <p>/complement(join(2594. 3341,3785. 3921,6539. 6804, 9264. 9420,9586. 9715))</p> <p>/gene="KLK9"</p> <p>/product="KLK9 mRNA"</p> <p>/complement(join(3192. 3341,3785. 3921,6539. 6804, 9264. 9420,9586. 9628))</p> <p>/gene="KLK9"</p> <p>/note="kallikrein-like protein 3 KLK-L3"</p> <p>/codon_start=1</p> <p>/product="kallikrein 9"</p> <p>/protein_id="AAG2255.1"</p> <p>/db_xref="GI:10799394"</p> <p>/translation="MKULGICALSLAGHWADTRALGAEGRPNSPQWAGLHFLT LSANCAITISRMVLTAAHCKPYLWRLGHEHLMKEGEQLFRTDFPFPENFD LIFNDHNDIMLRLPROARSPAVOPPLNLSQTVSGMCCIISGWGAVSPPLAPV TLGCAISILENKICMNAVPGHISDSMLGALGEGSGSCQDSGCELVNGLTAGV TGCAGPCSRPRPAVTVSVCHYDMIDIMEN"</p>
repeat_region	<p>3406. 3496</p> <p>/rpt_family="MSR1"</p>
repeat_region	<p>3558. 3700</p> <p>/rpt_family="MSR1"</p>
repeat_region	<p>4042. 4139</p> <p>/rpt_family="GA-rich"</p>
repeat_region	<p>4140. 4449</p> <p>/rpt_family="AluSx"</p>
repeat_region	<p>4450. 4483</p> <p>/rpt_family="GA-rich"</p>
repeat_region	<p>4529. 4812</p> <p>/rpt_family="AluSc"</p>
repeat_region	<p>complement(5043. 5253)</p> <p>/rpt_family="Char1168"</p>
repeat_region	<p>5370. 5480</p> <p>/rpt_family="AluO"</p>
repeat_region	<p>5605. 5712</p>





gene GSKERGVTVNVCRTYTNWIOKTIQAK  
complement (165420, .170283)  
/gene="KLK7"  
/note="synonym: SCCE"  
complement (join (165420, .165575, 167672, .167808, 168124, .168371, 169651, .169798, 170211, .170283))  
/gene="KLK7"  
/product="stratum corneum chymotryptic enzyme"  
complement (join (165420, .165575, 167672, .167808, 168124, .168371, 169651, .169798, 170211, .170283))  
/gene="KLK7"  
/note="serine protease"  
/codon\_start=1  
/product="stratum corneum chymotryptic enzyme"  
/protein\_id="AAG3360.1"  
/db\_xref="GI:11244765"  
/translation="MARSLPPLQIILSLALEAGEAGDKIINDAPCARSHPMO  
VALSGNOLHGVVAVNERVLTAAHCKMYTHVAGSDTIGBRRAORIKASVSPHP  
GYSOTVNDMLVKLNSQALSMVYKLPSCFPGTCTVSGGTTSDVYPP  
SDLMCVKXLSPODCTKYKDLLENMLCGLPDSKMAKNDGSGPLVCRGTGL  
VSWGTFPGQPDGVYTVCKFTKWIINDTKHR"  
complement (join (183943, .184098, 185635, .185768, 187865, .188127, 188293, .188452, 188957, .189036))  
/product="neutropsin"  
complement (join (183943, .184098, 185635, .185768, 187865, .188127, 188293, .188452, 188957, .189036))  
/note="serine protease; also called ovasin"  
/codon\_start=1  
/product="neutropsin"  
/protein\_id="AAG3361.1"  
/db\_xref="GI:11244766"  
/translation="NGRRPRPAKTMFLLLLGAGWAGSHAPQEDKVLGHECOPHSO  
FMQALFGQQLGGLVYVGNWVLTAAHCKMYTHVAGSDTIGBRRAORIKASVSPHP  
SIRPCTNSSPVEDHNDMLQRLDASGSKPYSIADHDSIQOKTYSKMTV  
TSPEPEPDLNCKAEVKIPEKSCEDAVPGYIDGMCASDKADTCQSDSGPLVC  
DGAIGITSWSDPGMSDKPQVYTNICRYLDIKIIGSG"  
complement (join (190980, .191129, 191573, .191709, 194324, .194589, 197048, .197204, 197370, .197412))  
/product="kallikrein-1-like 3"  
complement (join (190980, .191129, 191573, .191709, 194324, .194589, 197048, .197204, 197370, .197412))  
/note="serine protease"  
/codon\_start=1  
/product="kallikrein-1-like 3"  
/protein\_id="AAG3362.1"  
/db\_xref="GI:11244767"  
/translation="MKGLCALSLLAGHMAPTRAIGABECRPNQPMQALPHLT  
RLFCATLISRWLTLAHCERPYLWRLGEBHMKKEGEPQLFRYDPPHGNKD  
LSANDNDIMILIRPRARLSPAVQPNLSQTSVSPMOGLISGMKAVSPALRPV  
TLQCNMISILENKLCHWVPGHISDSMLCAGLWGGSGSCQGSGLVLCNGLAGV  
SGGAPCRPRPRPVYVSVCHYLDIMQIMEN"

Query Match 46.3%; Score 673.4; DB 9; Length 23000;  
Best Local Similarity 99.7%; Pred. No. 4, 1e-130;  
Matches 685; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 756 CCAGAGTGACTCTGAGAGCCCTGGTCTGTGACGAGACCCCTCCAGGCATCTCTGCTG 815  
DB 202820 CCAGAGTGACTCTGAGAGCCCTGGTCTGTGACGAGACCCCTCCAGGCATCTCTGCTG 202761

QY 816 GGGGTGTTAACCCCTGAGCTGTGACGAGATCCAGCTGTCTACACCCAGATCTTCATAA 875  
DB 202760 GGGGTGTTAACCCCTGAGCTGTGACGAGATCCAGCTGTCTACACCCAGATCTTCATAA 202701

QY 876 CATGCTCGAGTCAATTAAGTCATAGCTCCAGATGATGCTCAAGCTCCAGCTGA 935  
DB 202700 CATGCTCGAGTCAATTAAGTCATAGCTCCAGATGATGCTCAAGCTCCAGCTGA 202641

QY 936 TCAGAGTGTATGCTCCGCTGATCCAGATGCCAGAGCTCCATCGCTCACTCTCTCC 995  
DB 202640 TCAGAGTGTATGCTCCGCTGATCCAGATGCCAGAGCTCCATCGCTCACTCTCTCC 202581

QY 996 TCCCAAGTCGAGCTGAATCTCCCTGTGTCTGACATGTTCAAACTCTGCGCCCTCCACA 1055

DB 202580 TCCCAAGTCGAGCTGAATCTCCCTGTGTCTGACATGTTCAAACTCTGCGCCCTCCACA 202521

QY 1056 COTCAACATCTCCCTCTCACTCAATCTCCCAACATCTCCCACTCTCTGCTGACT 1115  
DB 202520 COTCAACATCTCCCTCTCACTCAATCTCCCAACATCTCCCACTCTCTGCTGACT 202461

QY 1116 GAGCTGAATGACAGAGAGTGTGCAAGAGTTTATTCAGAGAGCCAGAGCCGCTC 1175  
DB 202460 GAGCTGAATGACAGAGAGTGTGCAAGAGTTTATTCAGAGAGCCAGAGCCGCTC 202401

QY 1176 ATGACCCAGCTCTGAGAGGAGTTATCTGGGGTTCACCCCACTGACTCTGCACTCC 1235  
DB 202400 ATGACCCAGCTCTGAGAGGAGTTATCTGGGGTTCACCCCACTGACTCTGCACTCC 202341

QY 1236 CGCTGTGTGACTTTGGGCAAGCCAGAGTCCCTCTGACCTGATTTCTCTCATCTGCA 1295  
DB 202340 CGCTGTGTGACTTTGGGCAAGCCAGAGTCCCTCTGACCTGATTTCTCTCATCTGCA 202281

QY 1296 AATGGAACAATGACGTGCTTACTCTTTAGACATGTTGTGAGAGACTATGATTAACA 1355  
DB 202280 AATGGAACAATGACGTGCTTACTCTTTAGACATGTTGTGAGAGACTATGATTAACA 202221

QY 1356 TGTGTATGTAATCTTCTAT-GTGAATGTGATTAAGCTTAACAAGTGGTGTGAGTT 1414  
DB 202220 TGTGTATGTAATCTTCTATGTAATGTGATTAAGCTTAACAAGTGGTGTGAGTT 202161

QY 1415 CTGACTAAGGTTACTCTGTGTGCTGA 1441  
DB 202160 CTGACTAAGGTTACTCTGTGTGCTGA 202134

RESULT 10  
AC130782/c AC130782 200792 bp DNA linear HTG 22-NOV-2002  
LOCUS Pan troglodytes clone CH251-355A20, WORKING DRAFT SEQUENCE, 12  
DEFINITION Ordered pieces.  
AC130782  
VERSION AC130782.2 GI:25167101  
KEYWORDS HTG; HTGS PHASE2; HTGS\_DRAFT.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
REFERENCE  
AUTHORS Akheri N., Antonellis A., Ayele K., Beckstrom-Sternberg S.M.,  
Benjamin B., Blakesley R.W., Bouffard G.G., Brinkley C., Brooks S.,  
Cariaga K., Coleman B., Engle J., Granite S., Guan X., Gupta J.,  
Haghighi P., Han J., Hansen N., Ho S.-L., Idol J.R., Karlins B.,  
Laird P., Lee-Lin S.-Q., Legaspi R., Maduro Q.L., Maduro V.B.,  
Marques E.H., Mastello C., Maskeri B., McDowell J.,  
Paguirigan C., Pearson R., Portnoy M.E., Prasad A.,  
Reddix-Dugue N., Schandler K., Schueler M.G., Sison C.,  
Stattipop S., Thomas J.W., Thomas P.J., Touchman J.W., Vogt J.L.,  
Wetherby K.D., Wiggins L., Young A. and Green E.D.  
NISC Comparative Sequencing Initiative  
Unpublished  
2 (bases 1 to 200792)  
TITLE  
JOURNAL Green E.D.  
AUTHORS Direct Substitution  
TITLES Submitted (14-AUG-2002) NIH Intramural Sequencing Center, 8717  
JOURNAL Groveomont Circle, Gaithersburg, MD 20877, USA  
REFERENCE 3 (bases 1 to 200792)  
AUTHORS Green E.D.  
TITLES Direct Substitution  
JOURNAL Submitted (12-NOV-2002) NIH Intramural Sequencing Center, 8717  
Groveomont Circle, Gaithersburg, MD 20877, USA  
COMMENT On Nov 22, 2002 this sequence version replaced gi:22218452.  
-----  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc\_zoo@nigr.nih.gov

Db 133327 TGTGATGTAATCTTCATGGTGATTGTCATGTAAGCCTTAACACAGTGGGTGACTT 133268









BASE COUNT 110 a 183 c 113 g 126 t  
 ORIGIN  
 Query Match 35.0%; Score 508.4; DB 6; Length 532;  
 Best Local Similarity 99.4%; Pred. No. 1.7e-95;  
 Matches 531; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 712 AACACATGATATATGCTGAGACTGACACCGGGGCGAGACCTTGGCCAGATGACTGTGA 771  
 DB 1 AACACATGATATATGCTGAGACTGACACCGGGGCGAGACCTTGGCCAGATGACTGTGA 60  
 QY 772 GGGCCCCCTGCTGTGTGACGAGACCTCCAAAGGCATCTCTGTGGGGGTGTTTACCCCTGT 831  
 DB 61 GGGCCCCCTGCTGTGTGACGAGACCTCCAAAGGCATCTCTGTGGGGGTGTTTACCCCTGT 120  
 QY 832 GGCTGTGCCGAGATCCGAGCTGTCTACACCCAGATCTGAAATATATCTCTGATCAAT 891  
 DB 121 GGCTGTG-CGAGCATCCAGCTGTCTACACCCAGATCTGAAATATATCTCTGATCAAT 179  
 QY 892 AAAGTCATATGCTCCAGCTGATCCAGATGCTACGCTCAGCTGATCCAGATGTTATGCTC 951  
 DB 180 AAAGTCATATGCTCCAGCTGATCCAGATGCTACGCTCAGCTGATCCAGATGTTATGCTC 239  
 QY 952 CTGCTGATCCAGATCCAGAGGCTCCATCTGCTCACTCTTCTCCCAAGTGGCTGAA 1011  
 DB 240 CTGCTGATCCAGATCCAGAGGCTCCATCTGCTCACTCTTCTCCCAAGTGGCTGAA 299  
 QY 1012 CTCTCCCTTGTCTGACATGTTTAAACCTCTGCGCCCTCCACACCTTAACATCTCCC 1071  
 DB 300 CTCTCCCTTGTCTGACATGTTTAAACCTCTGCGCCCTCCACACCTTAACATCTCCC 359  
 QY 1072 CTCTGACCTGATCCCGGACCTTATCCAGATCTCTGCTGTACGAGCTGAATGCAAG 1131  
 DB 360 CTCTGACCTGATCCCGGACCTTATCCAGATCTCTGCTGTACGAGCTGAATGCAAG 419  
 QY 1132 AAGTGTGGCAAAAGTTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1191  
 DB 420 AAGTGTGGCAAAAGTTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 479  
 QY 1192 GAGCAGTTACTGGGGGTGACCCCAACCTGATCTCTGCGACTCCCGCTGTGTG 1245  
 DB 480 GAGCAGTTACTGGGGGTGACCCCAACCTGATCTCTGCGACTCCCGCTGTGTG 532

RESULT 14  
 LOCUS AX331572 532 bp DNA linear PAT 09-JAN-2002  
 DEFINITION Sequence 2081 from Patent WO0194629.  
 ACCESSION AX331572  
 VERSION AX331572.1 GI:18122206  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,  
 Horrigan, S., Soppet, D.R. and Weaver, Z.  
 Cancer gene determination and therapeutic screening using signature  
 gene sets  
 Patent: WO 0194629-A 2081 13-DEC-2001;  
 Avalon Pharmaceuticals (US)  
 Location/Qualifiers  
 1..532  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

FEATURES  
 source 1..532  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

BASE COUNT 110 a 183 c 113 g 126 t  
 ORIGIN  
 Query Match 35.0%; Score 508.4; DB 6; Length 532;  
 Best Local Similarity 99.4%; Pred. No. 1.7e-95;  
 Matches 531; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 712 AACACATGATATATGCTGAGACTGACACCGGGGCGAGACCTTGGCCAGATGACTGTGA 771  
 DB 1 AACACATGATATATGCTGAGACTGACACCGGGGCGAGACCTTGGCCAGATGACTGTGA 60  
 QY 772 GGGCCCCCTGCTGTGTGACGAGACCTCCAAAGGCATCTCTGTGGGGGTGTTTACCCCTGT 831  
 DB 61 GGGCCCCCTGCTGTGTGACGAGACCTCCAAAGGCATCTCTGTGGGGGTGTTTACCCCTGT 120  
 QY 832 GGCTGTGCCGAGATCCGAGCTGTCTACACCCAGATCTGAAATATATCTCTGATCAAT 891  
 DB 121 GGCTGTG-CGAGCATCCAGCTGTCTACACCCAGATCTGAAATATATCTCTGATCAAT 179  
 QY 892 AAAGTCATATGCTCCAGCTGATCCAGATGCTACGCTCAGCTGATCCAGATGTTATGCTC 951  
 DB 180 AAAGTCATATGCTCCAGCTGATCCAGATGCTACGCTCAGCTGATCCAGATGTTATGCTC 239  
 QY 952 CTGCTGATCCAGATCCAGAGGCTCCATCTGCTCACTCTTCTCCCAAGTGGCTGAA 1011  
 DB 240 CTGCTGATCCAGATCCAGAGGCTCCATCTGCTCACTCTTCTCCCAAGTGGCTGAA 299  
 QY 1012 CTCTCCCTTGTCTGACATGTTTAAACCTCTGCGCCCTCCACACCTTAACATCTCCC 1071  
 DB 300 CTCTCCCTTGTCTGACATGTTTAAACCTCTGCGCCCTCCACACCTTAACATCTCCC 359  
 QY 1072 CTCTGACCTGATCCCGGACCTTATCCAGATCTCTGCTGTACGAGCTGAATGCAAG 1131  
 DB 360 CTCTGACCTGATCCCGGACCTTATCCAGATCTCTGCTGTACGAGCTGAATGCAAG 419  
 QY 1132 AAGTGTGGCAAAAGTTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1191  
 DB 420 AAGTGTGGCAAAAGTTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 479  
 QY 1192 GAGCAGTTACTGGGGGTGACCCCAACCTGATCTCTGCGACTCCCGCTGTGTG 1245  
 DB 480 GAGCAGTTACTGGGGGTGACCCCAACCTGATCTCTGCGACTCCCGCTGTGTG 532

RESULT 15  
 LOCUS AX335105 532 bp DNA linear PAT 09-JAN-2002  
 DEFINITION Sequence 5614 from Patent WO0194629.  
 ACCESSION AX335105  
 VERSION AX335105.1 GI:18125824  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,  
 Horrigan, S., Soppet, D.R. and Weaver, Z.  
 Cancer gene determination and therapeutic screening using signature  
 gene sets  
 Patent: WO 0194629-A 5614 13-DEC-2001;  
 Avalon Pharmaceuticals (US)  
 Location/Qualifiers  
 1..532  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

FEATURES  
 source 1..532  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

BASE COUNT 110 a 183 c 113 g 126 t  
 ORIGIN  
 Query Match 35.0%; Score 508.4; DB 6; Length 532;  
 Best Local Similarity 99.4%; Pred. No. 1.7e-95;  
 Matches 531; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 712 AACACATGATATATGCTGAGACTGACACCGGGGCGAGACCTTGGCCAGATGACTGTGA 771  
 DB 1 AACACATGATATATGCTGAGACTGACACCGGGGCGAGACCTTGGCCAGATGACTGTGA 60  
 QY 772 GGGCCCCCTGCTGTGTGACGAGACCTCCAAAGGCATCTCTGTGGGGGTGTTTACCCCTGT 831  
 DB 61 GGGCCCCCTGCTGTGTGACGAGACCTCCAAAGGCATCTCTGTGGGGGTGTTTACCCCTGT 120



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 08:52:05 ; Search time 55 Seconds

(without alignments)  
1294.955 Million cell updates/sec

Title: us-10-021-368-1

Perfect score: 1496  
Sequence: 1 MRAPHLHLSAASGARALAK.....AVYQICKYMWINKVIRSN 276

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriopl:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1028	68.7	278	11 Q99W20	Q99W20 mus musculus
2	552.5	36.9	234	11 Q99W76	Q99W76 mus musculus
3	544	36.4	260	4 Q81M69	Q81M69 homo sapien
4	538	36.0	253	4 Q8N5N9	Q8N5N9 homo sapien
5	534	35.7	249	11 Q91YB3	Q91YB3 mus musculus
6	514	34.4	255	4 Q96R00	Q96R00 homo sapien
7	506	33.8	249	11 Q9QYN4	Q9QYN4 mus musculus
8	506	33.8	276	11 Q9QYN3	Q9QYN3 mus musculus
9	502	33.6	250	11 Q8CGR5	Q8CGR5 mus musculus
10	499.5	33.4	246	11 Q9R0T7	Q9R0T7 mus musculus
11	499.5	33.4	276	11 Q8CGR6	Q8CGR6 mus musculus
12	496	33.2	246	11 Q9QUR9	Q9QUR9 mus musculus
13	490	32.8	246	11 Q921R9	Q921R9 mus musculus
14	490	32.8	275	4 Q81XD7	Q81XD7 homo sapien
15	486	32.5	254	11 Q8CGR4	Q8CGR4 mus musculus
16	483	32.3	247	13 Q9W7Q5	Q9W7Q5 paratichthy

17	483	32.3	293	11 Q9D140	Q9D140 mus musculus
18	481	32.2	234	11 Q9R048	Q9R048 mus musculus
19	480.5	32.1	247	11 Q9CEN7	Q9CEN7 mus musculus
20	477.5	31.9	251	11 Q54854	Q54854 rattus norv
21	477	31.9	247	11 Q9D7Y7	Q9D7Y7 mus musculus
22	474.5	31.7	247	11 Q9CEN9	Q9CEN9 mus musculus
23	473.5	31.7	242	13 Q92099	Q92099 paratichthy
24	467.5	31.2	237	13 Q91515	Q91515 fugu rubrip
25	467.5	30.8	244	13 Q8Q5W3	Q8Q5W3 anguilla ja
26	461.5	30.8	246	11 Q88301	Q88301 mus musculus
27	461.5	30.8	253	11 Q91Y82	Q91Y82 mus musculus
28	461	30.8	239	4 Q8N1C9	Q8N1C9 homo sapien
29	461	30.8	241	13 Q987G9	Q987G9 engraulis j
30	458	30.6	251	4 Q8N2U3	Q8N2U3 homo sapien
31	454	30.3	240	13 Q987H0	Q987H0 engraulis j
32	454	30.3	247	13 Q42608	Q42608 petromyzon
33	453	30.3	247	13 Q42158	Q42158 petromyzon
34	452	30.2	238	13 Q9W7Q6	Q9W7Q6 paratichthy
35	450.5	30.1	222	13 Q8A7L1	Q8A7L1 oncorhynch
36	450.5	30.1	242	13 Q9W7Q7	Q9W7Q7 paratichthy
37	449.5	30.0	245	13 Q42160	Q42160 petromyzon
38	446.5	29.8	244	13 Q42159	Q42159 petromyzon
39	445.5	29.8	239	11 Q63275	Q63275 rattus norv
40	445.5	29.8	247	4 Q8NHW4	Q8NHW4 homo sapien
41	441.5	29.5	251	11 Q9DBQ8	Q9DBQ8 mus musculus
42	439.5	29.4	261	6 Q29474	Q29474 canis fami
43	438.5	29.3	235	11 Q63274	Q63274 rattus norv
44	438	29.3	249	13 Q9W6K0	Q9W6K0 notochenia
45	438	29.3	675	13 Q9W6J8	Q9W6J8 dissostichu

## ALIGNMENTS

RESULT 1  
Q99W20 PRELIMINARY; PRT; 278 AA.

AC Q99W20; 01-JUN-2001 (TRENKREI, 17, Created)  
DT 01-JUN-2001 (TRENKREI, 17, Last sequence update)  
DT 01-MAR-2003 (TRENKREI, 23, Last annotation update)  
DE Similar to kallikrein 10.  
GN 2300002A13RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN 11

RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
DR EMBL; BC002100; AA02100.1; -  
DR HSSP; P00763; IDPO.  
DR MEROPS; S01.084; -

DR MGD; MGI:1916790; 2300002A13RIK.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser-protease try.  
DR Pfam; PF00083; Trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00202; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
KW Hydrolase; Protease; Serine protease;  
SQ SEQUENCE 278 AA; 31043 MW; DD68B4AB9ABCCF55 CRC64;

Query Match 68.7%; Score 1028; DB 11; Length 278;

Best Local Similarity 68.5%; Pred. No. 7e+92;  
Matches 189; Conservative 33; Mismatches 52; Indels 2; Gaps 2;

QY 1 MRAPHLHLSAASGARALAK-LLPLMAQLMAAEALIPQNDTLDPAYAPCARSGQPW 59  
DB 1 MRVPLHLSAASGWSLVKLLPLMAQLMAAEALIPQNDTLDPAYAPCARSGQPW 60

```

QY 60 QVSLPENGISFHCAGVLDQSVWLTAAAC-GNKPLMARVGDHLLILQSGDLRTTTSVYH 118
DB 61 QVSLPENGISFHCAGVLDQSVWLTAAAC-GNKPLMARVGDHLLILQSGDLRTTTSVYH 120
QY 119 PKYHGSQSPILPRRDEHDLMLKLARPVVGPFRVALQPYRCAGPQDQVAGMGTGA 178
DB 121 PKYHGSQSPILPRRDEHDLMLKLARPVVGPFRVALQPYRCAGPQDQVAGMGTGA 180
QY 179 ARRYKYNKGLTCCSITTLISPKCEVEVPYGVVNTNMICAGLDKQDPCQSDSGPLVCDT 238
DB 181 SRRVYKYSLSGSKYVTLISQKQCEFPYGVITNSMICAEDKQDQSDSGPLVCDT 240
QY 239 LQGLISWGVPCSGAOPAVYTOICXKYSWINKVIR 274
DB 241 LQGLISWGVPCSGAOPAVYTOICXKYSWINKVIR 276

```

## RESULT 2

```

Q9CV76 PRELIMINARY; PRT; 234 AA.
AC Q9CV76;
DT 01-JUN-2001 (TRENBLREL. 17, Created)
DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
DE 2310008B01R1K protein (Fragment).
GN 2310008B01R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai T., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kasukawa T., Saito K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kankawa T., Saito K.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Mikado I., Pesole G., Quackenbush J.,
RA Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga M., Carinci P., de Bonaldo M.P.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guscinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Marzari I., Morbaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AK009217; BAB26143.1; -.
DR HSR; P00763; IDPO.
DR MEROPS; S01.020; -.
DR MGD; MGI:1916761; 2310008B01R1K.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00340; TRYPSIN_DOM; 1.
DR PROSITE; PS00344; TRYPSIN_HIS; 1.
DR PROSITE; PS00355; TRYPSIN_SER; 1.
KM Hydroxylase; Protease; Serine protease.
FT NON_TER 1
SQ SEQUENCE 234 AA; 25688 MW; 60815609EDD39110 CRC64;

```

Query Match

36.9%; score 552.5; DB 11; length 234;

```

Best Local Similarity 48.7%; Pred. No. 1,1e-45;
Matches 112; Conservative 29; Mismatches 80; Indels 9; Gaps 5;
QY 49 GAPCARSGPQVSLFENGISFHCAGVLDQSVWLTAAACGNKPLMARVGDHLLILQ-GE 107
DB 12 GVECKNSQSPQVQLFNGKYLRCQGVLDKRWLTAAHCKDV-VRLGHSILTKLDWTE 70
QY 108 QLRRTTSVHFKYHGSQSPILPRRDEHDLMLKLARPVVGPFRVALQPYRCAGPQD 167
DB 71 QLRRTTSITPSY-QGA-----YQNHEDRLRLRLRPIHLTRAVPVALPSSCVTTGA 124
QY 168 QCOVAGMCTTARARVKNKGLTCCSITTLISPKCEVEVPYGVVNTNMICAGLDKQDPCQ 227
DB 125 MCHYSKMTTKKPPDPDRDLQCLANLSTVENCTRAFPFRVETNMLCAGGKADKQCG 164
QY 228 DSGGPLVCDTLLQGLISWG-VYPCSGAOPAVYTOICXKYSWINKVIRSN 276
DB 185 DSGGPLVCGVLDGLVSMGVSVPQSGKIGVYTKVCKYTDWIRIVIRN 234

```

## RESULT 3

```

Q81W69 PRELIMINARY; PRT; 260 AA.
AC Q81W69;
DT 01-MAR-2003 (TRENBLREL. 23, Created)
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
DE 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DT Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strussberg P.;
RL Submitted (JPC-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC040867; AAH40867.1; -.
KM Hypothetical protein.
SQ SEQUENCE 260 AA; 28090 MW; EF5934EB96295660 CRC64;

```

Query Match

36.4%; score 544; DB 4; length 260;

Best Local Similarity 43.7%; Pred. No. 8.5e-45;

Matches 111; Conservative 27; Mismatches 104; Indels 12; Gaps 4;

```

QY 23 LMAQLMAABALLPQNDRLDPRAYGAPCARSGPQVSLFENGISFHCAGVLDQSVW 82
DB 16 LLLGAMVAGHSR--AQEDKVLG---GHECQPHSQPQALFQGOQLCGVLDQSVW 69
QY 83 TAARCGNKLPAARVGDHLLILQ-EGLRRTTSVHFKYHGSQSPILPRRDEHDLML 141
DB 70 TAARCKKRYVRLGDSHSLQKQGPQGLIVGSIPIPCINSDD-----VEDNHDLML 124
QY 142 KLARPVVGPFRVALQPYRCAGPQDQVAGMGTGAARVKNKGLTCCSITTLSPKEC 201
DB 125 QLRDQASLGSVKYKIDSLADHCTQPGKCTVSGMGTIVSPENPDTLNCLEVIIFPQKC 184
QY 202 EYVYGVVNTNMICAGLDKQDPCQSDSGPLVCDTLLQGLISWGVPCSGAOPAVYTO 261
DB 185 EDAYFGQITDVAVAGSSKADITQSDSGPLVCDGLQGLITWSPDPCRSKXPGYTN 244
QY 262 ICKYMSWINKVIR 275
DB 245 ICRYLDWKITIGS 258

```

## RESULT 4

```

Q8NSN9 PRELIMINARY; PRT; 253 AA.
AC Q8NSN9;
DT 01-OCT-2002 (TRENBLREL. 22, Created)
DT 01-OCT-2002 (TRENBLREL. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)

```

DB Kallikrein 7 (chymotrypsin, stratum corneum).  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NCBI\_TaxID=9606;

RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RA Strauberg R.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR EMBL: BC032005; AA032005.1; -;  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PRO0722; CHYMOTRYPSIN.  
 DR SMART: SM00200; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 253 AA; 27608 MW; 2D68BA41B22A668 CRC64;

Query Match 36.0%; Score 538; DB 4; Length 253;  
 Best Local Similarity 40.5%; Pred. No. 3, 1e-44;  
 Matches 107; Conservative 45; Mismatches 98; Indels 14; Gaps 4;

QY 14 ARALAKLPVLAQAQMAAEALLPQNDRLDPEAVAPCARSGQPMOVSLENGLSFHCG 73  
 DB 2 ARSLPLPLQILLSLAETGSEBQDQKID---GAPCARSHHPQVALLSGNQLHCG 57  
 QY 74 VLVDQSVVLTAAHCGNKPFLMARVGDHLLLOGEQLRTTSHVHKYKQSGPIIPRT 133  
 DB 58 VLVNERVLTAAHCGNKPFLMARVGDHLLLOGEQLRTTSHVHKYKQSGPIIPRT 108  
 QY 134 DEHDMLKLARVVEGPRRALQIPYRCAGQDQGVAGKGTAAARKYKNGLTGCSI 193  
 DB 109 HVDNDMLVLRDEVEKSSKVEAVQLPEHCEPPTGCTVSQGMGTTTSPDVFPSDLMCSOV 168  
 QY 194 TILSPKECEVFEPGVVNNMTCAGL--DRGQDPCQSDSGPLVCDETLQGLISNGVYPCGS 252  
 DB 169 KLTSSRCCKVYNDLQGLKMTLCAIGIPDSKNTCNDSGSEPLVGNPDLQGLVSGVTPCGQ 228  
 QY 253 AQHRAVYTOICKYMINKYIRSN 276  
 DB 229 PNDGVTYQVCKFKTKWINDTKKH 252

## RESULT 5

QY 091VE3 PRELIMINARY; PRT; 249 AA.

AC 091VE3;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Thymopsin (stratum corneum chymotrypsin enzyme) (Kallikrein 7)  
 DE (chymotrypsin, stratum corneum).  
 GN KLT7.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Thymus;  
 RA Yamaguchi N.;  
 RT "A novel cDNA cloning of mouse serine protease, thymopsin, i.";  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/Svj;  
 RA Hansson L., Backman A., Ny A., Edlund M., Edholm E., Tornell J.,  
 Wallbrandt P., Egeirud T.;

RT "Epidermal overexpression of stratum corneum chymotrypsin enzyme in  
 RT mice; a model for chronic itchy dermatitis";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.

RA Strauberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=2354683; PubMed=12466851;  
 RA The FANTOM Consortium  
 RA the Riken Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573 (2002).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR EMBL: AB008371; BAB5604.1; -;  
 DR EMBL: AF339930; AAK69652.1; -;  
 DR EMBL: BC027823; AA027823.1; -;  
 DR EMBL: AK029477; BAC26467.1; -;  
 DR EMBL: AK077406; BAC36787.1; -;  
 DR HSP; P00761; IAN1.  
 DR MGD; MGI:1346336; KLT7.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PRO0722; CHYMOTRYPSIN.  
 DR PROSITE: PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 249 AA; 27257 MW; 0D4E380F12D14F87 CRC64;

Query Match 35.7%; Score 534; DB 11; Length 249;  
 Best Local Similarity 40.2%; Pred. No. 7, 6e-44;  
 Matches 106; Conservative 45; Mismatches 89; Indels 24; Gaps 7;

QY 17 LAKLPLPMAQMAAEALLPQNDRLDPEAVAPCARSGQPMOVSLENGLSFHCGVLY 76  
 DB 5 LSLITVLS--LALTA--GQSERIID---GYKESHPQVALLSGNQLHCGVLY 56  
 QY 77 DQSVLTAAHCGNKPFLMARVGDHLLLOGEQLR--RTTSVHPRKYNQSGPIIPRT 133  
 DB 57 DKVWLTAAHCGNKGQVQVQSGDKI---GDQAQKIKATKSPRHGQY-----STXT 104  
 QY 134 DEHDMLKLARVVEGPRRALQIPYRCAGQDQGVAGKGTAAARKYKNGLTGCSI 193  
 DB 105 HVDNDMLVLRDEVEKSSKVEAVQLPEHCEPPTGCTVSQGMGTTTSPDVFPSDLMCSOV 164  
 QY 194 TILSPKECEVFEPGVVNNMTCAGL--DRGQDPCQSDSGPLVCDETLQGLISNGVYPCGS 252  
 DB 165 KLTSSRCCKVYNDLQGLKMTLCAIGIPDSKNTCNDSGSEPLVGNPDLQGLVSGVTPCGQ 224  
 QY 253 AQHRAVYTOICKYMINKYIRSN 276  
 DB 225 PNDGVTYQVCKFKTKWINDTKKH 248

## RESULT 6

QY 096RQ0 PRELIMINARY; PRT; 255 AA.

AC 096RQ0;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Prostogen.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

EX MEDLINE=2126193; PubMed=11327827;  
 RA Takayama T.K., Carter C.A., Deng T.;  
 RT "Activation of prostate-specific antigen precursor (pro-PSA) by  
 RT prostatic, a novel human prostatic serine protease identified by  
 RT degenerate PCR";  
 RL Biochemistry 40:1679-1687(2001).  
 CC -1- SIMILARITY: BELONGS TO PERTIDASE FAMILY S1.  
 DR EMBL, AF303046; AAK62813.1; -.  
 DR HSSP; P00761; 1ANI.  
 DR MEROPS; S01.081; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser. protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYP\_SPC; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 KW Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 255 AA; 2796 MW; 00D5B79E14B9469F CRC64;  
 Query Match 34.4%; Score 514; DB 4; Length 255;  
 Best Local Similarity 41.2%; Pred. No. 6, 3e-42;  
 Matches 99; Conservative 41; Mismatches 80; Indels 20; Gaps 4;  
 QY 49 GAPCARSGPQVQVSLFNGSLFHCAGVLDQSWVLTAAHCGNKPLMARVGDHLLLLQGE 107  
 DB 24 GDECAFPHSQVQVVALYERGFNCASLISPHWVLSAHCQSRFMRVRLGSHNLRKDGFE 83  
 QY 108 QLRRTTSVHPKHYQSGPILPRITDEHMLKLARPVVPGPRVALQLPYCAQPGD 167  
 DB 84 QLRRTTSVHPKHYQSGPILPRITDEHMLKLARPVVPGPRVALQLPYCAQPGD 135  
 QY 168 QCVVAGMGTTA-----ARRVKNKGITCGSITLSPKCEVEYFGVVTNNMICAG 217  
 DB 136 ACVVSQMGVLSHNEPFGTAGSPRSQVSLPDLHCANIGILISDSCDKSEYPRILTNTWCAG 195  
 QY 218 LD-AGQPCGSDSGPILVCDLTLGGIISNGVYPCGSAQHPAVYTOICKYSWINKVIRSN 276  
 DB 196 AEGRAESCEGDSGPVLCGGLIISNGVYPCGSAQHPAVYTOICKYSWINKVIRSN 255  
 RESULT 7  
 Q9QYN4 PRELIMINARY; PRT; 249 AA.  
 AC Q9QYN4;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE HIPOSTASIN (2310015108RIK protein).  
 GN PRS820 OR 2310015108RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Yamaguchi N.; Mitsui S.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Tongue;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Baralov S., Casavant T.,  
 RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,  
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Funuo M., Aono H., Baldarelli R., Barh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaudo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guernicich S., Hill D., Hofmann M., Hune D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Komdaert P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,  
 RA Wyszynski B., Yoshida K., Hasegawa Y., Kawanji H., Kohetsuki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 DR EMBL; AB016226; BAA88825.1; -.  
 DR EMBL; AK009320; BAB26461.1; -.  
 DR EMBL; AK009360; BAB26241.1; -.  
 DR HSSP; P00763; IDBO.  
 DR MEROPS; S01.257; -.  
 DR MGD; MGI:192977; PRS820.  
 DR InterPro; IPR001254; Ser. protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR SMART; SM00020; TRYP\_SPC; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 249 AA; 27604 MW; P9FP9CB45D7D2725 CRC64;  
 Query Match 33.8%; Score 506; DB 11; Length 249;  
 Best Local Similarity 41.3%; Pred. No. 4e-41;  
 Matches 95; Conservative 44; Mismatches 85; Indels 6; Gaps 3;  
 QY 49 GAPCARSGPQVQVSLFNGSLFHCAGVLDQSWVLTAAHCGNKPLMARVGDHLLLLQGE 107  
 DB 24 GDECAFPHSQVQVVALYERGFNCASLISPHWVLSAHCQSRFMRVRLGSHNLRKDGFE 83  
 QY 108 QLRRTTSVHPKHYQSGPILPRITDEHMLKLARPVVPGPRVALQLPYCAQPGD 167  
 DB 84 QLRRTTSVHPKHYQSGPILPRITDEHMLKLARPVVPGPRVALQLPYCAQPGD 139  
 QY 168 QCVVAGMGTTAARRVKNKGITCGSITLSPKCEVEYFGVVTNNMICAGLDR-GQPCQ 226  
 DB 140 SLLISGWTSSPQRLPRLPHSLRCANVSLIEKCEKAPGNITDTMCAVFKGKSCQ 199  
 QY 227 SPGGPILVCDLTLGGIISNGVYPCGSAQHPAVYTOICKYSWINKVIRSN 276  
 DB 200 GDSGPELVNCSLGGIISNGVYPCGSAQHPAVYTOICKYSWINKVIRSN 249  
 RESULT 8  
 Q9QYN3 PRELIMINARY; PRT; 276 AA.  
 AC Q9QYN3;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE HIPOSTASIN prostate type.  
 GN PRS820.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN RP SEQUENCE FROM N.A.  
 RC TISSUE=Prostate;  
 RA Yamaguchi N.; Mitsui S.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Prostate;



RX MEDLINE=20525460; PubMed=11072088;  
 RA Mitsun S., Okui A., Kominami K., Uemura H., Yamaguchi N.;  
 RT "DNA cloning and tissue-specific splicing variants of mouse  
 h1pocastain/Tisp (PRSS20).";  
 RL Biochim. Biophys. Acta 1494:206-210(2000).  
 CC - - SIMILARITY: BELONGS TO PEPTIDASE FAMILY SL.  
 DR EMBL; AB016227; BA336955.1; -.  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.257; -.  
 DR MG1; MGI.1929977; Prss20.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser.protease\_Try.  
 DR Pfam; PF00089; Trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRY\_ SPC; 1.  
 DR PROSITE; PSS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 276 AA; 30753 MW; 90BDC0A8AB178D6 CRC64;

Query Match 33.8%; Score 506; DB 11; Length 276;  
 Best local Similarity 41.3%; Pred. No. 4.5e-41;  
 Matches 95; Conservative 44; Mismatches 85; Indels 6; Gaps 3;

QY 49 GAFQANGSGPMOYSLFNGSLFHCAGVLDQSWLTAAHCKNKLMAVGDHLLLLQG-E 107  
 DB 51 GVECRPHSQPMQVALFOKTRLLCGATLLAPKMLTFAHCKRPYVILLSEHMLEKTDGE 110  
 QY 108 QLRRTSRVYHPKHYHSGSPILPRRDEHMLIKLARVVGPRRALQLYRCQRPD 167  
 DB 111 QRRMATESPHDPENNS---LNNKRRNDIMLVKSSVFFTRAQPTLLSPHCVAAGT 166  
 QY 168 QCVAGWGTAAARVRYKNGLTCSSITLLSPKCEVFYGVVYNNMICAGLDR-GQDFCQ 226  
 DB 167 SCLISGMGTSSPQLRPHSLRCANVSIIEHKECEKAYGNITDMLCGSVKKEGDSQ 226  
 QY 227 SSGGFLVCDFTLQGLTSGVYPCGSAQHPRAYTQICKMSNTINKYRN 276  
 DB 227 GSGGFLVCDFTLQGLTSGVYPCGSAQHPRAYTQICKMSNTINKYRN 276

RESULT 9  
 Q8CGR5 PRELIMINARY; PRT; 250 AA.

AC Q8CGR5; 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Glandular kallikrein KLK14.  
 GN KLK14.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=23235484; PubMed=12437987;  
 RT Olsson A.Y., Lundvall A.;  
 RT "Organization and evolution of the glandular kallikrein locus in Mus  
 musculus.";  
 RL Biochem. Biophys. Res. Commun. 299:305-311(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Adams M., Mural R.;  
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AY152433; AAN78421.1; -.  
 SQ SEQUENCE 250 AA; 27016 MW; F6FEBF2290FEB8 CRC64;

Query Match 33.6%; Score 502; DB 11; Length 250;  
 Best local Similarity 39.8%; Pred. No. 9.9e-41;  
 Matches 104; Conservative 43; Mismatches 96; Indels 18; Gaps 6;

QY 20 LFLPLMAQLMAEAPALLPQNDRLDPEAVGAPCARSGPMOYSLFNGSL--FHCAGVLD 77  
 DB 4 LLLIIIOA---LAVAIAGSGDKHI---IGYICVNSQPMQVALQAGPHEFLCGVLLS 57  
 QY 78 QSWVTAAHCKNKLPMARVGDHLLLLQG-EQLRRTSRVYHPKHYHSGSPILPRRDEH 136  
 DB 58 DQWVITFAHCKARPIHLVALGKINIRMENTQGVNVARQPHRPQYR-----QAHN 109  
 QY 137 DMLIKLARPVVPGRRVALQLPYCAQPGDCCQVAGMCTTAARVRYKNGLTCSSITLL 196  
 DB 110 DMLIKLQKRVLGAVKXISVASSCASPGPCRVSGMGTIASPIARPTALQCVNVNIM 169  
 QY 197 SPECEVFYGVVYNNMICAGL-DRGDPCCGSDSGSPVCDFTLQGLTSGVYPCGSAH 255  
 DB 170 SQGACHRAVFGIITSGMVCAGVPEGKQSDSGSPVCGQLQGLTSGVSWMERCAHPY 229  
 QY 256 PAVYTOICKMSWINKYRN 276  
 DB 230 PGVYANLGVYHSWIDRTWQSN 250

RESULT 10  
 Q9ROT7 PRELIMINARY; PRT; 246 AA.

AC Q9ROT7; 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Pancreatic trypsin (0910001B19RIK protein) (Trypsinogen 8).  
 GN TD OR 0910001B19RIK OR TRYPSINOGEN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129SVJ;  
 RX MEDLINE=99436155; PubMed=10506205;  
 RA Ohmura K., Kohno N., Kobayashi Y., Yamagata K., Sato S.,  
 RA Kaishibara S., Baba T.,  
 RT "A homologue of pancreatic trypsin is localized in the acrosome of  
 mammalian sperm and is released during acrosome reaction.";  
 RL J. Biol. Chem. 274:29426-29432(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Stomach, and Spleen;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai U., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bares G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bolt C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guttingich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima M., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Rongwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,  
 RA Wyszewski A., Yoshida K., Hasegawa Y., Kawaji H., Kontecki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.

RC STRAIN=129;  
 RX MEDLINE=21103195; PubMed=11160223;  
 RA Chen F., Rowen B., Hood L., Rothenberg E.V.,  
 RT "Differential transcriptional regulation of individual TCR Vbeta

RT segments before gene rearrangement";  
 CC J. Immunol. 166:1771-1780(2001).  
 DR EMBL; AB017032; BAB74761.1; -  
 DR EMBL; AK008667; BAB25821.1; -  
 DR EMBL; AK003064; BAB22542.1; -  
 DR EMBL; AB000664; BAB69056.1; -  
 DR HSSP; P00763; ISLU.  
 DR MEROPS; S01.057; -  
 DR MGD; MG11913350; 0910001B19R1K.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser protease\_try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYPSIN\_DOM; 1.  
 DR PROSITE; PSS0240; TRYPsin\_DOM; 1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 KM Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 246 AA; 26274 MW; B6A3F4C9079633F CRC64;

Query Match 33.4%; Score 499.5; DB 11; Length 246;  
 Best Local Similarity 40.9%; Pred. No. 1.7e-40;  
 Matches 108; Conservative 41; Mismatches 94; Indels 21; Gaps 8;

QY 15 RALAKLPLIMADLMAEAALLPQNDRLDPEAYGAPCARQSPQVSLFNGLSFHCAGY 74  
 DB 2 RAL-----LFLALVGAVA--FPVDD--DKIVGVTGCRNSVPYQVLSNGYHF CGGS 51  
 QY 75 LVDSQWTLTAHGCNPKPLARVGDHLLIQG-EQLKRTTSVYVPHKTHGSGPILPRRT 133  
 DB 52 LINDQWVVAHACHYKSRIQVRLGHEHNNVLEGNQFVNSKIIKHNFNS-----RT 103  
 QY 134 DEHDLMILKLARPVSPFRALQLPYRCAPQDQCVAGMGTAAARVYVNGKLTCSGI 193  
 DB 104 LNDNDIMIKIASVYTNARVATVALPSSCAPAGTQCLISGNGTILSPGVNPDLLQCLDA 163  
 QY 194 TITSPKECEVFPQVNTNNMICAG-LDRGDPQCSGSGPLVCDETLQGLISGVYPCGS 252  
 DB 164 PLTPQADCEASYPGKTNMNICVGFLEGGHDSQCSGSGPVCNGQLQGLIVSWG-YGCAL 222  
 QY 253 AQHNAVYTOICKYMSWINKYIRSN 276  
 DB 223 KDNPGVTKVCNVDNIQNTIAAN 246

## RESULT 11

Q8CGR6 PRELIMINARY; PRT; 276 AA.

AC Q8CGR6; 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Glandular kallikrein KLK13.  
 GN KLK13.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22325484; PubMed=12437987;  
 RA Olsson A.Y., Lundwall A.;  
 RT "Organization and evolution of the glandular kallikrein locus in Mus musculus.";  
 RL Biochem. Biophys. Res. Commun. 299:305-311(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Adams N., Mural R.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; A1152432; AAN78420.1; -  
 SQ SEQUENCE 276 AA; 30360 MW; 33E189C67492FDCA CRC64;  
 RT

Query Match 33.4%; Score 499.5; DB 11; Length 276;  
 Best Local Similarity 43.5%; Pred. No. 1.9e-40;  
 Matches 101; Conservative 34; Mismatches 86; Indels 11; Gaps 6;

## RESULT 12

Q90UK9 PRELIMINARY; PRT; 246 AA.

AC Q90UK9; 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 23, Last annotation update)  
 DE TESP4 (0910001B19R1K protein) (Trypsinogen 9).  
 GN TC OR 0910001B19R1K OR TRYPSINOGEN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN NCBI\_TaxID=10090;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=99436155; PubMed=10506205;  
 RA Ohmura K., Kohno N., Kobayashi Y., Yamagata K., Sato S.,  
 RA Kashwabara S., Baba T.;  
 RT "A homologue of pancreatic trypsin is localized in the acrosome of mammalian sperm and is released during acrosome reaction.";  
 RL J. Biol. Chem. 274:29426-29432(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=pancreas;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hata A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gasteirland T., Gissi C., King B., Kochiava H.,  
 RA Kuehl P., Lewis S., Matsumoto Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Mashio T.,  
 RA Sakai K., Oikawa T., Furuno M., Aono H., Balarelli R., Barsh G.,  
 RA Blake J., Boileau D., Bolunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Nombaretti P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=21103195; PubMed=11160223;  
 RA Chen F., Rowen L., Hood L., Rothenberg E.V.;  
 RT "Differential transcriptional regulation of individual TCR Vbeta

RT segments before gene rearrangement.";  
 RL J. Immunol. 166:1771-1780(2001).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR EMBL; AB009661; BAA85187.1; -  
 DR EMBL; AB017031; BAA74760.1; -  
 DR EMBL; AK007843; BAB25100.1; -  
 DR EMBL; AB000664; AAB69088.1; -  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.058; -  
 DR MGD; MG1:191350; 0910001B19R1K.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser. protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYP\_SEC; 1.  
 DR PROSITE; PS50240; TRYPSEIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSEIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSEIN\_SER; 1.  
 KW Hydrolyase; Protease; Serine protease.  
 SO SEQUENCE 246 AA; 2627 MM; 915C92353EC42809 CRC64;

Query Match 33.2%; Score 496; DB 11; Length 246;  
 Best Local Similarity 41.0%; Pred. No. 3.7e-40;  
 Matches 105; Conservative 41; Mismatches 94; Indels 16; Gaps 7;

QY 23 LMAQLMAAFAALLPQNDRLDPEAYGAPCARSGOPMVSLEFNGLSFHCAGVLDQSWYL 82  
 DB 5 LFLALVGAFAVA--FPVDD--DKIVGGYTCRENSVYQVSLNSGNYH-CGSSLINDQWV 59  
 QY 83 TAAHCGKPKMARVGDHLLILQSG-EQLRRTTRSVVHPKRYHGSGLPRLPRDHDMLL 141  
 DB 60 SAHACYRTRIQVRLGHNINVLGNEQFVNSAKIKHPNPN-----RKLINDIML 111  
 QY 142 KLRPVVPGPRVVALOLPYRCAPQDQCQVAGGTAAARVKNKNGLTCSITLSPKRC 201  
 DB 112 KLASPVTLNARVATVALPSSCAPAGTCCLSGKNTLSFGVNNPDLLQCLDAPLDAQDC 171  
 QY 202 EEPYGVVNTNMICAG-LDRGDPCCQSDSGPLVCDLTLGLISWGVYPCGSAQHPAVYT 260  
 DB 172 EASYPGKITNMICAGFLBGGKDSQQDSGSPVVCNGELQGLISWG-YGCALPDNGVYT 230  
 QY 261 QICKYMSINKVIRSN 276  
 DB 231 KVCNIVDMIDDTIAAN 246

## RESULT 13

ID Q921R9 PRELIMINARY; PRT; 246 AA.

AC Q921R9 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Trypsinogen 16.  
 GN TRYGN16 OR TRYPSINOGEN.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Balb/c;  
 RA "Rowen L., Hood L.;  
 RT "Comparison between strains Balb/c and 129 in a region of the mouse T  
 cell receptor beta locus."  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Balb/c;  
 RX MEDLINE=21101195; PubMed=11160223;  
 RA Chen F., Rowen L., Hood L., Rothenberg E.V.;  
 RT "Differential transcriptional regulation of individual TCR beta  
 segments before gene rearrangement.";

RL J. Immunol. 166:1771-1780(2001).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR EMBL; AF107342; AAC79093.1; -  
 DR EMBL; AB000665; AAB69088.1; -  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.063; -  
 DR MGD; MG1:2148749; Trypsin16.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser. protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYP\_SEC; 1.  
 DR PROSITE; PS50240; TRYPSEIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSEIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSEIN\_SER; 1.  
 KW Hydrolyase; Protease; Serine protease.  
 SO SEQUENCE 246 AA; 26134 MM; 34E173B18CA2F463 CRC64;

Query Match 32.8%; Score 490; DB 11; Length 246;  
 Best Local Similarity 40.2%; Pred. No. 1.4e-39;  
 Matches 103; Conservative 43; Mismatches 94; Indels 16; Gaps 7;

QY 23 LMAQLMAAFAALLPQNDRLDPEAYGAPCARSGOPMVSLEFNGLSFHCAGVLDQSWYL 82  
 DB 5 LFLALVGAFAVA--FPVDD--DKIVGGYTCRENSVYQVSLNSGNYH-CGSSLINDQWV 59  
 QY 83 TAAHCGKPKMARVGDHLLILQSG-EQLRRTTRSVVHPKRYHGSGLPRLPRDHDMLL 141  
 DB 60 SAHACYRTRIQVRLGHNINVLGNEQFVNSAKIKHPNPN-----RKLINDIML 111  
 QY 142 KLRPVVPGPRVVALOLPYRCAPQDQCQVAGGTAAARVKNKNGLTCSITLSPKRC 201  
 DB 112 KLASPVTLNARVATVALPSSCAPAGTCCLSGKNTLSFGVNNPDLLQCLDAPLDAQDC 171  
 QY 202 EEPYGVVNTNMICAG-LDRGDPCCQSDSGPLVCDLTLGLISWGVYPCGSAQHPAVYT 260  
 DB 172 EASYPGKITNMICAGFLBGGKDSQQDSGSPVVCNGELQGLISWG-YGCALPDNGVYT 230  
 QY 261 QICKYMSINKVIRSN 276  
 DB 231 KVCNIVDMIDDTIAAN 246

## RESULT 14

ID Q81XD7 PRELIMINARY; PRT; 275 AA.

AC Q81XD7 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Variant form of hypostasin/CLK1.  
 GN CLK1.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue=Prostate;  
 RA Nakamura T., Mitsui S., Miki T., Yamaguchi N.;  
 RT "Molecular cloning and expression of a variant form of  
 hypostasin/CLK1 in prostate."  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB078780; BAC54105.1; -  
 SO SEQUENCE 275 AA; 30165 MM; 257A42B28F40E2C4 CRC64;

Query Match 32.8%; Score 490; DB 4; Length 275;  
 Best Local Similarity 36.2%; Pred. No. 1.6e-39;  
 Matches 104; Conservative 50; Mismatches 91; Indels 42; Gaps 8;

QY 19 KLPDLMAQLMAAFAALLPQNDRLDPEAYGAPCARSGOPMVSLEFNGLSFHCAGVLDQ 78  
 DB 2 RIQQLTLLAL---AGLVGGETRI--IKGECKPHSQPMQALREKTRRLQAGATLIV 54

```
QY 79 SWLTAHCGNKPLM-----ARVGDHLLLLQG-EOLR 110
D 55 RWLTLAAHC-LKP-WVSLTSPHVSPLSSSNVCLSHLSRYIVHLGQHNLQKEGCEQTR 112
QY 111 RTTSVVRKXKHQSGPLPRTRDEHMLKLARPVVGGRRVALQLPFCAPQDQC 170
D 113 TATESFPHPGNNS---LEPKDHRNDIMLVKMASPVSTWAVRPLTLSSRCVAGTSTCL 168
QY 171 VAGGTTAARRKYNKGLTSSITLSPKECEVFYPGVVTNNMICAGL-DRGDPQCQSDS 229
D 169 ISGSGSTSSPQLRPHLTRCANITILIEKQCEANYPGNITIDTWVCASVQEGKDCQSDS 228
QY 230 GGPVVCETLQGLISKGVYPCGSAQHPAVTTOICKYMSINKVIRSN 276
D 229 GGPVVCNOSLOGIISWGQDPCALTRKPGVYTKVCKYVDWMIQETMKNN 275
```

## RESULT 15

```
Q8CGR4 PRELIMINARY; PRT; 254 AA.
AC Q8CGR4;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Prostln.
GN KUK15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=2325484; Pubmed=12437987;
RA Olsson A.Y., Lundwall A.;
RT "Organization and evolution of the glandular kallikrein locus in Mus
RT musculus."
RL Biochem. Biophys. Res. Commun. 299:305-311(2002).
[2]
RE SEQUENCE FROM N.A.
RA Adams M., Mural R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY152434; AN78422.1;
SQ SEQUENCE 254 AA; 28042 MW; AA9E38BEBDD01861 CRC64;
```

Query Match 32.5%; Score 486; DB 11; Length 254;  
Best Local Similarity 40.2%; Pred. No. 3.6e-39;  
Matches 106; Conservative 38; Mismatches 94; Indels 26; Gaps 7;

```
QY 23 LMAQLMAAARALLPQNDTRLDEAYGAPCARSGOPQVSLFNGLSFHCGAVLVDSWYL 82
D 3 LLLAFVLVSA--QGDKVLB-----GEECVHSQPMQVALFERGRFCAPLISPRVYL 56
QY 83 TAHHCGNKPLMAVGDHLLLLQG-EOLRRTTSVVRKXKHQSGPLPRTRDEHMLL 141
D 57 TAHCCTRRMRVVLGHNLRKRFDPQLRSVSRILPHPGYE-----AFTHRHIMLL 108
QY 142 KLARPVVPGRRVALQLPFCAPQDQCQVAGW-----GTTAAR--VKYNGGLTCS 191
D 109 RLFKPARLTAIVRPVALPRCPILIGDCVAVSGMLSDNNPAGATGSKSHVRLPTLHCA 168
QY 192 STTILSPKECEVFYPGVVTNNMICAGL-DRGDPQCSDSGPLVCDETLOGLISKGVYPC 250
D 169 NISIISEASCNKDYPERVLPTWVCAGVEGGGTDSGSDGSPVLCGALQGIIVSGVDVPC 228
QY 251 GSAQHPAVTTOICKYMSINKVIR 274
D 229 DYTTRKGVYTKVCKYVDWMIQETMKNR 252
```

Search completed: November 25, 2003, 09:08:42  
Job time : 60 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 07:31:39 ; Search time 17 Seconds

(without alignments)  
763.492 Million cell updates/sec

Title: US-10-021-368-1

Perfect score: 1496  
1 MRAPHLHLSAGAPALAKL.....AVTIOCKYMKHINKVIRSN 276

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	1496	100.0	276	1	KLKA_HUMAN
2	564	37.7	248	1	KLKC_HUMAN
3	549	36.7	253	1	KLK7_HUMAN
4	547	36.6	260	1	KLK6_HUMAN
5	520	34.8	256	1	KLK6_HUMAN
6	513	34.3	260	1	NREP_MOUSE
7	506	33.8	247	1	TRY2_CANFA
8	505.5	33.8	247	1	TRY2_BOVIN
9	503.5	33.7	250	1	KLK6_HUMAN
10	502	33.6	250	1	KLK6_HUMAN
11	500	33.4	251	1	KLK6_HUMAN
12	499.5	33.4	260	1	NREP_MOUSE
13	497.5	33.3	277	1	KLK6_HUMAN
14	497	33.2	243	1	TRY1_BOVIN
15	494	33.0	246	1	TRY2_MOUSE
16	488.5	32.7	246	1	TRY2_MOUSE
17	488.5	32.7	246	1	KLK6_HUMAN
18	486.5	32.5	247	1	TRY1_MOUSE
19	483	32.3	247	1	TRY1_MOUSE
20	482	32.2	246	1	TRY1_MOUSE
21	481	32.2	244	1	KLK6_HUMAN
22	477.5	31.9	247	1	TRY3_MOUSE
23	470	31.4	247	1	TRY2_MOUSE
24	469.5	31.4	247	1	TRY2_MOUSE
25	468.5	31.3	248	1	TRY3_MOUSE
26	468.5	31.3	248	1	TRY3_MOUSE
27	467	31.2	244	1	TRY2_MOUSE
28	464.5	31.0	246	1	TRY2_MOUSE
29	460	30.7	246	1	TRY1_MOUSE
30	458.5	30.6	247	1	TRY1_MOUSE
31	457.5	30.5	261	1	KLK7_MOUSE
32	457	30.5	243	1	TRY1_MOUSE
33	452	30.2	248	1	TRY1_MOUSE

34	452	30.2	248	1	TRY2_CHICK	Q90628 gallus galli
35	452	30.2	260	1	ESTR_CANFA	P09582 canis familiaris
36	450.5	30.1	231	1	TRY2_SALSA	P35032 salmo salar
37	445.5	29.8	242	1	TRY1_SALSA	P35031 salmo salar
38	443	29.6	241	1	TRY1_GADMO	P16049 gadus morhua
39	436	29.1	241	1	TRY1_GADMO	P16049 gadus morhua
40	436	29.1	259	1	KLK6_MOUSE	P00757 mus musculus
41	434.5	29.0	256	1	KLK4_MOUSE	P36375 mus musculus
42	431.5	28.8	244	1	KLK6_MOUSE	P00757 mus musculus
43	429	28.7	261	1	KLK6_MOUSE	P36375 mus musculus
44	427.5	28.6	261	1	KLK6_MOUSE	P36375 mus musculus
45	426.5	28.5	259	1	KLK6_MOUSE	P07647 rattus norvegicus

## ALIGNMENTS

RESULT 1  
KLKA\_HUMAN STANDARD; PRT; 276 AA.  
ID: KLKA\_HUMAN  
AC: O43240; Q99920; Q9G2M9;  
DT: 15-DEC-1998 (Rel. 37, Created)  
DT: 15-DEC-1998 (Rel. 37, Last sequence update)  
DT: 15-SEP-2003 (Rel. 42, Last annotation update)  
DE: Kallikrein 10 precursor (EC 3.4.21.-) (Protease serine-like 1) (Normal epithelial cell-specific 1).  
GN: KLK10 OR PRKSL1 OR NES1.  
OS: Homo sapiens (Human).  
OC: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
RP: SEQUENCE FROM N.A.  
RC: TISSUE=Epithelium;  
MEDLINE=96320486; PubMed=8764136;  
RA: Liu X.-L., Wazer D.E., Matanabe K., Band V.  
RT: "Identification of a novel serine protease-like gene, the expression of which is down-regulated during breast cancer progression."  
RL: Cancer Res. 56:3371-3379(1996).  
[2]  
RP: SEQUENCE FROM N.A.  
RC: MEDLINE=96321170; PubMed=9647736;  
RA: Luo L.-Y., Heidrick J.A., Scherer S.W., Beatty B., Squire J., Diamandis E.P.  
RT: "Structural characterization and mapping of the normal epithelial cell-specific 1 gene."  
RL: Biochem. Biophys. Res. Commun. 247:580-586(1998).  
[3]  
RP: SEQUENCE FROM N.A.  
RC: MEDLINE=20510030; PubMed=11054574;  
RA: Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J., Moss P., Paepker B., Wang K.  
RT: "Sequencing and expression analysis of the serine protease gene cluster located in chromosome 19q13 region."  
RL: Gene 257:119-130(2000).  
[4]  
RP: SEQUENCE FROM N.A.  
RC: Lamedin J.E., McCready P.M., Skowronski B., Viswanathan V., Panharth-Schulz K., Gordon L., Dias J., Ramirez M., Stilaagen S., Buxton H., Velasco N., Do L., Regala W., Terry A., Brower A., Gaines J., Dangane L., Erler A., Christensen M., Georgescu A., Avila J., Liu S., Andrease T., Trankheim M., Attix C., Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B., Atellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.  
RT: "Sequence analysis of chromosome 19q13.4."  
RL: Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.  
[5]  
RP: SEQUENCE FROM N.A.  
RC: TISSUE=Uterus;  
MEDLINE=22388257; PubMed=12477932;  
RA: Strassberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish R.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stjepic M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer T.E.,  
 RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raza S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallary S.J.,  
 RA Roark S.A., McGowan P.J., McKernan K.J., Malek J.A., Gurnatone P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Parley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting R., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.M., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RP CHARACTERIZATION. PubMed=9809976;  
 RX MEDLINE=99025848; PubMed=9809976;  
 RA Goyal J., Smith K.M., Cowan J.M., Wazer D.E., Lee S.W., Band V.;  
 RT "The role for NSE1 serine protease as a novel tumor suppressor".  
 RL Cancer Res. 58:4782-4786(1998).  
 CC -1- FUNCTION: HAS A TUMOR-SUPPRESSOR ROLE FOR NSE1 IN BREAST AND  
 CC PROSTATE CANCER.  
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BREAST, OVARY AND PROSTATE.  
 CC -1- DEVELOPMENTAL STAGE: DOWN-REGULATED DURING BREAST CANCER  
 CC PROGRESSION.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. KALLIKREIN SUBFAMILY.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AF024605; AAB1602.1; -  
 DR EMBL; AF055461; AAC14266.1; -  
 DR EMBL; AF243527; AAC33363.1; -  
 DR EMBL; AF011473; AAC33256.1; -  
 DR EMBL; BC002710; AAH02710.1; -  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.246; -  
 DR Genew; HGNC:6358; KLK10.  
 DR MIM; 602673; -  
 DR GO; GO:0005576; C:extracellular; TMS.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYP\_SEC; 1.  
 DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE; PS00135; TRYPsin\_SER; FALSE NEG.  
 KW Hydroxylase, Serine protease, Glycoprotein, Signal, Anti-oncogene.  
 FT SIGNAL 1 30  
 FT CHAIN 31 276  
 FT ACT\_SITE 86 86  
 FT ACT\_SITE 137 137  
 FT ACT\_SITE 229 229  
 FT DISULFID 52 162  
 FT DISULFID 71 87  
 FT DISULFID ? 263  
 FT DISULFID 169 235  
 FT DISULFID 201 215  
 FT DISULFID 225 250  
 FT CARBOHYD 39 50  
 FT CONFLICT 50 50  
 FT CONFLICT 149 149  
 P -> L (IN REF. 3 AND 4).

SEQ SEQUENCE 276 AA; 30138 MW; 82A2507379BAB313 CRC64;  
 Query Match 100.0%; Score 1496; DB 1; Length 276;  
 Best Local Similarity 100.0%; Pred. No. 2,1e-128;  
 Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRAPHILHSAAAGSRALAKLPLILMAQLMAABALPONTFRDPDEYAGAPCARGSQPMQ 60  
 DB 1 MRAPHILHSAAAGSRALAKLPLILMAQLMAABALPONTFRDPDEYAGAPCARGSQPMQ 60  
 QY 61 VSLFNGSLFHCAGVLVDQSVWLTAAHGNKPLMARVGDHLLLOGBQLRTTRSVVHPK 120  
 DB 61 VSLFNGSLFHCAGVLVDQSVWLTAAHGNKPLMARVGDHLLLOGBQLRTTRSVVHPK 120  
 QY 121 YHSGSGPLPRTRTEHMLKLARPVPRPRALDLPFCRQPGQCCVAGAGTTAAR 180  
 DB 121 YHSGSGPLPRTRTEHMLKLARPVPRPRALDLPFCRQPGQCCVAGAGTTAAR 180  
 QY 181 RVKYNKGLTCSITILSPKECEVPYGVVNTNMIAGLDHGDGPPCSGSGPLVCDETLQ 240  
 DB 181 RVKYNKGLTCSITILSPKECEVPYGVVNTNMIAGLDHGDGPPCSGSGPLVCDETLQ 240  
 QY 241 GILSGVYPCGSAQHAPVYVYQICRYMNTNKVIRSN 276  
 DB 241 GILSGVYPCGSAQHAPVYVYQICRYMNTNKVIRSN 276  
 RESULT 2  
 KKC HUMAN STANDARD; PRT; 248 AA.  
 ID KKC HUMAN STANDARD; PRT; 248 AA.  
 AC Q9URK0; Q9URK1; -  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Kallikrein 12 precursor (BC 3.4.21.-) (Kallikrein-like protein 5)  
 DE (KUK-15).  
 GN KLK12 OR KLK15.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euteleostomi; Primates; Carnivora; Hominoidea; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RX SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=20118156; PubMed=10652563;  
 RA Yousef G.M., Luo L.-Y., Diamandis E.P.;  
 RT "Identification of novel human kallikrein-like genes on chromosome  
 RT 19q13.3-q13.4".  
 RL Anticancer Res. 19:2843-2852(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RA Yousef G.M., Maglana A., Scorilas A., Diamandis E.P.;  
 RT "Cloning of new alternatively spliced forms of the kallikrein-like  
 RT gene 5 (KUK-15)".  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=20510030; PubMed=11054574;  
 RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,  
 RA Moss P., Paepker B., Wang K.;  
 RT "Sequencing and expression analysis of the serine protease gene  
 RT cluster located in chromosome 19q13 region".  
 RL Gene 257:119-130(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RA Lamerding J.E., McCreedy P.M., Skowronski E., Viswanathan V.,  
 RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillsen S.,  
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Gaines J.,  
 RA Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,  
 RA Andreise T., Frankheim M., Attix C., Amico-Keller G., Coeffield B.,  
 RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,  
 RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,  
 RA Olsen A.S., Carraro A.V.;  
 RT "Sequence analysis of chromosome 19q13.4".

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9UKR0-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9UKR0-2; Sequence=VSP\_005403;  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. KALLIKREIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF135025; AAD26426.2; -;  
 DR EMBL, AF135025; AAR06065.1; -;  
 DR EMBL, AF135025; AAG33365.1; -;  
 DR EMBL, AC011473; AAG23258.1; -;  
 DR HSSP, P00763; IDPO.  
 DR MEROP, S01\_020; -;  
 DR Genew; HGNC:6360; KLIK2.  
 DR MIM: 605539; -;  
 DR GO: 0005576; C:extracellular; NAS.  
 DR GO: 0006508; P:proteolysis and peptidolysis; NAS.  
 DR InterPro: IPR001254; Ser\_protease\_fam.  
 DR Pfam: PF00089; Trypsin; 1.  
 DR SMART; SM00020; TRYP\_SPCL; 1.  
 DR PROSITE; PS00240; TRYPSPIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSPIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSPIN\_SER; 1.  
 KM Hydrolyase; Serine protease; Glycoprotein; signal;  
 KM Alternative splicing.  
 FT SIGNAL 17  
 FT CHAIN 18 248  
 FT ACT\_SITE 62 62 KALLIKREIN 12.  
 FT ACT\_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 28 161 BY SIMILARITY.  
 FT DISULFID 47 63 BY SIMILARITY.  
 FT DISULFID 133 235 BY SIMILARITY.  
 FT DISULFID 140 206 BY SIMILARITY.  
 FT DISULFID 172 186 BY SIMILARITY.  
 FT DISULFID 196 222 BY SIMILARITY.  
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 236 248 KYVDWTRMIMRN -> NSTLVGLSTSNRNSQPF (in  
 FT isoform 2).  
 SQ SEQUENCE 248 AA; 26733 MW; BBA73B98F8BAF703 CRC64;  
 Query Match 37.7%; Score 564; DB 1; Length 248;  
 Best Local Similarity 45.9%; Pred. No. 5.4e-44;  
 Matches 118; Conservative 29; Mismatches 92; Indels 18; Gaps 5;  
 QY 23 LHAQMAARALPQNDTRUDPBAI-GAPCARGSPPQVSLPNCISHTCAGVLDQSNV 81  
 Db 7 LLLCVGLSLQDA-----TPKINGTECGNSQPMQVGFESTSLRCGGVLLDHRKV 57  
 QY 82 LTAAGCNKPLMARVGDHLLDQ-GEQRTTRSVVHPKQSGSGLIPRTDEHML 140  
 Db 58 LTAAGGSGSYWRRLGHSHLSQLDWTEQIRHGSFVTHGYLGAS-----TSHHDLRL 111  
 QY 141 LKARPRVSPRYALQPRCKQSDQCVAGNGTAAARVKNKGLTSSITLTSKE 200  
 Db 112 LKRLPRVTSVQPPPLPMDCAVTAETCHVSGWGTNHRNPFPLLQCLNLSVSHAT 171  
 QY 201 CEVFPVGVNMMICAGLDKRGDPCQSDSGSPLVCDTELQGLISWG-VYPCSSAGHPAVY 259  
 Db 172 CHGVYPRITSNMVCAGGVGQDACCQSDSGSPLVCGVTLQGLVMSGVGPGCGDGIPIGVY 231

QY 260 TOICKMYSWINKVIRSN 276  
 Db 232 TYICKYVMIMIMRN 248  
 RESULT 3  
 ID KLIK7 HUMAN STANDARD; PRI; 253 AA.  
 AC P49862;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic  
 DE enzyme) (hSCCE).  
 GN KLIK7 OR PRS56 OR SCCE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.  
 RC TISSUE=Skin;  
 RX MEDLINE=94308225; PubMed=8034709;  
 RA Hansson L., Stromqvist M., Backman A., Wallbrant P., Carlstein A.,  
 RA Egelund T.;  
 RT "Cloning, expression, and characterization of stratum corneum  
 RT chymotryptic enzyme. A skin-specific human serine proteinase.";  
 RL J. Biol. Chem. 269:19420-19426(1994).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Keratinocytes;  
 RA Yousef G.M., Scortias A., Diamandis E.P.;  
 RT "Molecular characterization, mapping and tissue expression of the  
 RT human stratum corneum chymotryptic enzyme gene";  
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20510030; PubMed=11054574;  
 RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,  
 RA Moss P., Paepel B., Wang K.;  
 RT "Sequencing and expression analysis of the serine protease gene  
 RT cluster located in chromosome 19q13 region.";  
 RL Gene 257:119-130(2000).  
 RN (4)  
 RP SEQUENCE FROM N.A.  
 RA Hansson L., Backman A., Ny A., Edlund M., Edholm E., Tornell J.,  
 RA Wallbrant P., Egelund T.;  
 RT "Epidermal overexpression of stratum corneum chymotryptic enzyme in  
 RT mice; a model for chronic itchy dermatitis";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN (5)  
 RP CHARACTERIZATION.  
 RX MEDLINE=95314630; PubMed=7794273;  
 RA Skvrt A., Stromqvist M., Egelund T.;  
 RT "Primary substrate specificity of recombinant human stratum corneum  
 RT chymotryptic enzyme";  
 RL Biochem. Biophys. Res. Commun. 211:586-589(1995).  
 CC -1- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE  
 CC STRUCTURES IN THE CORNEATED LAYER OF THE SKIN IN THE CONTINUOUS  
 CC SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID  
 CC RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE  
 CC CLEAVES INSULIN B CHAIN AT 6-LEU--CYS-7, 16-TYR-1-LEU-17, 25-PHE-  
 CC -TYR-26, AND 26-TYR-1-THR-27. COULD PLAY A ROLE IN THE ACTIVATION  
 CC OF PRECURSORS TO INFLAMMATORY CYTOKINES.  
 CC -1- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS  
 CC EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE  
 CC ALSO SEEN IN THE BRAIN AND KIDNEY.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. KALLIKREIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L33404; AAC7551.1; -  
DR EMBL: AF166330; AAD49718.1; -  
DR EMBL: AF243527; AAG33360.1; -  
DR EMBL: AF332583; AAK69624.1; -  
DR PIR: A53968; A53968.  
DR HSSP: P00763; IDPO.  
DR MEROPS: S01.300; -  
DR Genew; HGNC:6368; KLK7.  
DR MIM: 604438; -  
DR GO: GO:0008544; P:epidermal differentiation; TAS.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser protease\_Try.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PRO0022; CHYMOTRYPSIN.  
DR SMART: SM00020; TRYPSIN\_DOM; 1.  
DR PROSITE: PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE: PS00134; TRYPSIN\_SIS; 1.  
DR PROSITE: PS00135; TRYPSIN\_SIS; 1.  
KM Hydroxylase; Serine protease; Zymogen; Glycoprotein; Signal.  
FT SIGNAL 1 22  
FT PROPEP 23 29  
FT CHAIN 30 233  
FT ACT\_SITE 70 70  
FT ACT\_SITE 112 112  
FT ACT\_SITE 205 205  
FT DISULFID 36 137  
FT DISULFID 55 71  
FT DISULFID ? 239  
FT DISULFID 144 211  
FT DISULFID 176 190  
FT DISULFID 201 226  
FT CARBOHYD 246 246  
SQ SEQUENCE 253 AA; 27525 MW; 2D68B6515A76A668 CRC64;

Query Match 36.7%; Score 548; DB 1; Length 253;  
Best Local Similarity 40.9%; Pred. No. 1,36-42;  
Matches 108; Conservative 45; Mismatches 97; Indels 14; Gaps 4;

QY 14 ABAALAKLPPLMAQMAAEALLPONDRLDPEAVGAPARSGQWQVSLFNGLSFRCAG 73  
DB 2 ASLSLLPLDILLSTALSTLSTGEAEAGCDKIID-----GAPCARQSHPMQVALLSGNOLHCGG 57  
QY 74 VLVDSQSVLTAAHCGNKPIMAEVGDHLLLLQGEQLRTTSVHPKXYHOGSGFILPRT 133  
DB 58 VLVNERWVLTAAHCKMNEVYTHLGSDDLGDRAQRI-KAKSKFRHPGV-----STQT 108  
QY 134 DEHDMLLKLARPVVPGPRVRLQPRYCAQPGDCCVAGWGTTPAARVTKNKGLTCSI 193  
DB 109 HYNMLLVLTNSQARLSMVKVRLPSRCEPGTCTVSGWGTTSPPVTSPSDIMCYDV 168  
QY 194 TILSPKECEVEYGVVNTNMICAGL-DRGQDPCQSDSGEPLVCDFTLGGISMGVYPCGS 252  
DB 169 KLISQDCTKVKYKDLBLSMGLCAGIPDSKXKACNDSGEPVLCRTLGGISLVSGTFPCGQ 228  
QY 253 AOHNAVYQICKYMINKVIRSN 276  
DB 229 PNDPGVYTVQVCKFKTKWINDYMKKH 252

RESULT 4  
KLK8 HUMAN STANDARD; PRT; 260 AA.  
AC 060259; Q9HCB3; Q9U1D9; Q9UQ47;  
BT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DR Neuropein precursor (EC 3.4.21.-) (NP) (Kalikrein 8) (Ovasin) (Serine  
protease TADG-14) (Tumor-associated differentially expressed gene-14

DE protein).  
GN KLK8 OR PRSS19 OR TADG14 OR NRPN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteleota; Primates; Carnivora; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Hippocampus;  
RX MEDLINE=98372070; PubMed=9714609;  
RA Yoshida S., Taniguchi M., Hirata A., Shiosaka S.;  
RT "Sequence analysis and expression of human neuropein cDNA and gene";  
RL Gene 213:9-16(1998).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC TISSUE=Brain;  
RX MEDLINE=99203457; PubMed=10102990;  
RA Mitsu S., Tsuruoka N., Yamashiro K., Nakazato H., Yamaguchi N.;  
RT "A novel form of human neuropein, a brain-related serine protease, is  
RT generated by alternative splicing and is expressed preferentially in  
RL human adult brain";  
RN Eur. J. Biochem. 260:627-634(1999).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Ovary;  
RX MEDLINE=99413504; PubMed=10485494;  
RA Underwood L.J., Tamimoto H., Wang Y., Shigemasa K., Parmlay T.H.,  
O'Brien T.J.;  
RT "Cloning of tumor-associated differentially expressed gene-14, a novel  
RL serine protease overexpressed by ovarian carcinoma";  
RN Cancer Res. 59:4435-4439(1999).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Gan L., Gelinas R., Gown A.M., Moss P., Smith R., Wang K.;  
RT "Molecular cloning and characterization of a novel serine protease,  
RT ovasin, a potential molecular marker for ovarian carcinomas";  
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=20510030; PubMed=11054574;  
RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,  
Moss P., Paepfer B., Wang K.;  
RT "Sequencing and expression analysis of the serine protease gene  
RT cluster located in chromosome 19q13 region";  
RL Gene 257:119-130(2000).  
RN [6]  
RP SEQUENCE OF 1-164 FROM N.A. (ISOFORM 1).  
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,  
Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,  
Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Barnes  
Dangahan L., Exler A., Christensen M., Georgescu A., Avila J., Liu S.,  
Andrease T., Trankheim M., Attix C., Amico-Keller G., Coffield J.,  
Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kromtiller B.,  
Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,  
Olsen A.S., Carrano A.V.;  
RT "Sequence analysis of chromosome 19q13.4";  
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.  
RN [7]  
RP FUNCTION: SUGGESTED TO BE INVOLVED IN KINDLING EPILEPTOGENESIS AND  
CC HIPPOCAMPAL PLASTICITY.  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-, Lys-.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=O60259-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=O60259-2; Sequence=VSP\_005401;  
CC -1- TISSUE SPECIFICITY: ISOFORM 1 IS PREDOMINANTLY EXPRESSED IN THE  
CC PANCREAS WHILE ISOFORM 2 IS EXPRESSED IN ADULT BRAIN AND  
CC HIPPOCAMPUS. BOTH FORMS ARE ALSO FOUND IN FETAL BRAIN AND  
CC PLACENTA. NOT DETECTED IN KIDNEY, SPLEEN, LIVER AND LUNG.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. KALIKREIN SUBFAMILY.





DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PS00722; CHYMOTRYPSIN.  
 DR PROSITE: PS50240; TRYPSIN\_DOM. 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS. 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER. FALSE NEG.  
 KM Hydroxylase: Serine protease; Glycoprotein; Signal; Zymogen;  
 KM Alternative splicing.  
 FT SIGNAL 1 16  
 FT PROPEP 17 21  
 FT CHAIN 22 256  
 FT ACT\_SITE 62 62  
 FT ACT\_SITE 106 106  
 FT ACT\_SITE 209 209  
 FT CARBOHD 171 171  
 FT CARBOHD 232 232  
 FT VASPLIC 122 206  
 FT VASPLIC 122 256  
 FT VASPLIC 161 161  
 FT VASPLIC 162 256  
 FT CONFLICT 147 160  
 FT SEQUENCE 256 AA; 28087 MM; B5BFB9D6022786B5 CRC64;  
 SQ  
 Query March 34.8%; Score 520; DB 1; Length 256;  
 Best local similarity 39.9%; Pred. No. 5.4e-40;  
 Matches 107; Conservative 46; Mismatches 89; Indels 26; Gaps 6;  
 QY 21 LPLMAQIWAEEALLPQNDRLDPEAYGAPCARSGSPQVSLFNGLSFRCAGVLVDQSW 80  
 DB 3 LLLTSLSEFLASTAA--DQGDYLDE---GDECAHSPQPMVVALYERGRFNGCGLSLSPHW 56  
 QY 81 VLTAHCGNCKELMAYVGDHLLDQ--EQLRTSRVYHKYQSGSPILPRRTDHDLM 139  
 DB 57 VLSNAHCSRMWRVRLSEHNLRKDGPEQLRTSRVYHKYQSGSPILPRRTDHDLM 108  
 QY 140 LKLTARVPVPPRYVALQLPYRCAQPGDQCVAGMGTTA-----ARRVYKNGKLT 189  
 DB 109 LLRLVQFARLNPQVRPAVLFRCEHPGACVSGMGLVSHNBPCTAGSPRQVSLDPLTH 168  
 QY 190 CSSTIISPRKECFYPPGVVNNMTCAGLD--RGDDPQSGSGPLVCDLQGITLSNGVY 248  
 DB 169 CANSTISIDTSCDMSYPERLNTNVCAGABERGASGSDSGPLVCGGTLQGITLSNGVY 228  
 QY 249 PGGAQHPAVYVYTOICKYMSWINKYIRSN 276  
 DB 229 FCDNTTKPGVYTKVCHYLEMIRETKRN 256  
 RESULT 6  
 NREP\_MOUSE STANDARD; PRT; 260 AA.  
 AC Q61955;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neuropilin precursor (EC 3.4.21.-) (NP) (Kallikrein 8).  
 GN KLR8 OR PRSS19 OR NRPN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/c; TISSUE=Hippocampus;  
 RX MEDLINE=95348617; PubMed=7623137;  
 RA Chen Z.-L., Yoshida S., Kato K., Momota Y., Suzuki J., Tanaka T.,  
 RA Ito J., Mishino H., Aimoto S., Kiyama H., Shiosaka S.;  
 RT "Expression and activity-dependent changes of a novel limbic-serine  
 RT

RT protease gene in the hippocampus.";  
 RT U. Neurosci. 15:5088-5097(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yoshida S., Hirata A., Inoue N., Shiosaka S.;  
 RT "Cloning and assignment of mouse neuropilin gene, Prs19 to chromosome  
 RT 7B4.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF N-TERMINUS, AND CHARACTERIZATION.  
 RC STRAIN-BALB/c; TISSUE=Brain;  
 RX MEDLINE=98225202; PubMed=9556608;  
 RA Shimizu C., Yoshida S., Shibata M., Kato K., Momota Y., Matsumoto K.,  
 RA Shiosaka T., Mochizuki R., Kamachi T., Kawabe A., Shiosaka S.;  
 RT "Characterization of recombinant and brain neuropilin, a  
 RT plasticity-related serine protease.";  
 RL J. Biol. Chem. 273:11169-11196(1998).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 31-257.  
 RC TISSUE=Hippocampus;  
 RX MEDLINE=99134351; PubMed=9933620;  
 RA Kishi T., Kato M., Shimizu T., Kato K., Matsumoto K., Yoshida S.,  
 RA Shiosaka S., Hoshikawa T.;  
 RT "Crystal structure of neuropilin, a hippocampal protease involved in  
 RT kindling epileptogenesis.";  
 RL J. Biol. Chem. 274:4220-4224(1999).  
 CC -1- FUNCTION: SUGGESTED TO BE INVOLVED IN KINDLING EPILEPTOGENESIS AND  
 CC HIPPOCAMPAL PLASTICITY. HAS A STRONG PROTEOLYTIC ACTIVITY AGAINST  
 CC FIBRONECTIN  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-, Lys-  
 CC -1- ENZYME REGULATION: STRONGLY INHIBITED BY DIISOPROPYL  
 CC FLUOROPHOSPHATE, LEUPEPTIN AND (4-AMIDINO)BENZYL METHANESULFONYL 1-  
 CC FLUORIDE.  
 CC -1- SUBCELLULAR LOCATION: secreted.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE LIMBIC SYSTEM OF  
 CC MOUSE BRAIN AND IS LOCALIZED AT HIGHEST CONCENTRATION IN PYRAMIDAL  
 CC NEURONS OF THE HIPPOCAMPAL CA1-3 SUBFIELDS.  
 CC -1- MASS SPECTROMETRY: MW=26613; METHOD=MALDI; RANGE=29-260.  
 CC -1- MASS SPECTROMETRY: MW=26229; METHOD=MALDI; RANGE=33-260.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. KALLIKREIN SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 CC EMBL, D30785; BA06451.1; -  
 CC EMBL, AB032202; BA02435.1; -  
 CC PIR, I56559; I56559.  
 CC PDB, INPW; 2J-WAR-99.  
 CC MEROPS, S01.244; -  
 CC MGD; MGI:892018; KLR8.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PS00722; CHYMOTRYPSIN.  
 DR SMART, SMO0020; TRYP\_SPC. 1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM. 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS. 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER. 1.  
 KM Hydroxylase: Serine protease; Glycoprotein; Zymogen; Signal;  
 KM 3D-structure.  
 FT SIGNAL 1 28  
 FT PROPEP 29 32  
 FT CHAIN 33 260  
 FT ACT\_SITE 73 73  
 FT ACT\_SITE 120 120  
 FT ACT\_SITE 212 212  
 FT DISTLFD 39 173  
 FT DISULFID 58 74  
 FT

## N-LINKED (GLCNAC. . .) (POTENTIAL).

FT DISULFID 145 246  
 FT DISULFID 152 218  
 FT DISULFID 184 198  
 FT DISULFID 208 233  
 FT CARBOHYD 110 110  
 FT STRAND 34 34  
 FT STRAND 37 37  
 FT STRAND 41 42  
 FT STRAND 45 46  
 FT STRAND 47 52  
 FT STRAND 53 54  
 FT STRAND 55 64  
 FT STRAND 65 66  
 FT STRAND 67 70  
 FT HELIX 72 74  
 FT STRAND 80 83  
 FT STRAND 87 87  
 FT STRAND 88 89  
 FT STRAND 96 98  
 FT STRAND 100 105  
 FT STRAND 107 108  
 FT STRAND 114 115  
 FT STRAND 118 119  
 FT STRAND 122 126  
 FT STRAND 140 141  
 FT STRAND 148 149  
 FT STRAND 151 156  
 FT STRAND 170 170  
 FT STRAND 172 178  
 FT HELIX 181 187  
 FT STRAND 189 191  
 FT STRAND 194 195  
 FT STRAND 200 203  
 FT STRAND 202 203  
 FT STRAND 206 206  
 FT STRAND 209 210  
 FT STRAND 212 213  
 FT STRAND 215 218  
 FT STRAND 219 220  
 FT STRAND 221 228  
 FT STRAND 235 235  
 FT STRAND 236 237  
 FT STRAND 238 238  
 FT STRAND 240 244  
 FT HELIX 245 256  
 SQ SEQUENCE 260 AA; BESF6F6BE37CD608 CRC64;

Query Match 34.3%; Score 513; DB 1; Length 260;  
 Best Local Similarity 41.6%; Pred. No. 2, 4e-39;  
 Matches 106; Conservative 32; Mismatches 105; Indels 12; Gaps 5;

QY 20 LPELMAQMAEALLPQNDTRLDPEAYGAFCAFGSQPMQVSLFNGLSFHCAGVLYDQS 79  
 DB 13 ILLLEMGAWA--GLTRAGSKILE---GRECIHSPQMPAAUFGQERLLCGGVLYGDR 66  
 QY 80 WILTAAGCNKMLMRYVDDHLL--QGEOLARTRSVYVHKYHSGSPILPRTRDEHL 138  
 DB 67 WILTAAGCNKQKQYVRLDHSLOSRODPEQELQVAGSIQHPCIN--NSNP---EDSHSDY 121  
 QY 139 MLKTAAPVPGPRVRAQLPYKCAQPGDQCVAGWGTAAARVYKNGKLTCCSITLLSP 198  
 DB 122 MLIRLQNSANLGDKYVPQVLANLCPRVQKCIISGMGTVSPOEINFPTLNCARVLTYSQ 181  
 QY 199 KEEVFPYGVVNNMTCAGLDGDPCCSDSGSPVCELTGSLTSGVYVCGAAGHPV 258  
 DB 182 NKCBRYAPKITEGAVCGSSNGADTCCGDSGGLVCGMGLGITSWSDPCGKPERXGV 241  
 QY 259 YTOICKYMSINKVI 273  
 DB 242 YTKICKYTWIKKTM 256

RESULT 7

TRV2\_CANFA  
 ID TRV2\_CANFA STANDARD; PRT; 247 AA.  
 AC P06872;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trypsin, anionic precursor (EC 3.4.21.4).  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Flesipedida; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86284628; PubMed=3841794;  
 RA Plinsky S.D., Laforge K.S., Scheele G.;  
 RT "Differential regulation of trypsinogen mRNA translation: full-length  
 RT mRNA sequences encoding two oppositely charged trypsinogen isoenzymes  
 RT in the dog pancreas."  
 RL Mol. Cell Biol. 5:2659-2676 (1985).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg|-Xaa, Lys|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M1589; AAA30899.1; -  
 DR PIR; A25273; TRDG.  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.258; -  
 DR Interpro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; Trypsin; 1.  
 DR SMART; SM00020; TRYP\_Spc; 1.  
 DR PROSITE; PS00240; TRYP\_SIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYP\_SIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYP\_SIN\_SER; 1.  
 KW Hydrolyase, serine protease; Digestion; Pancreas; Zymogen; Signal;  
 KM Multigene family.  
 FT SIGNAL 1 15  
 FT PROPEP 16 23  
 FT CHAIN 24 247  
 FT ACT\_SITE 63 63  
 FT ACT\_SITE 107 107  
 FT ACT\_SITE 200 200  
 FT DISULFID 30 160  
 FT DISULFID 48 64  
 FT DISULFID 132 233  
 FT DISULFID 139 206  
 FT DISULFID 171 185  
 FT DISULFID 196 220  
 FT SITE 194 194  
 SQ SEQUENCE 247 AA; 26423 KM; 374E9D1DEDBEAF CRC64;

Query Match 33.8%; Score 506; DB 1; Length 247;  
 Best Local Similarity 41.8%; Pred. No. 9, 7e-39;  
 Matches 107; Conservative 41; Mismatches 92; Indels 16; Gaps 8;

QY 23 LMAQMAEALLPQNDTRLDPEAYGAFCAFGSQPMQVSLFNGLSFHCAGVLYDQSVL 82  
 DB 5 LIAFLGAAVAR--PTDD--DKIVGGTCENSVPYQVSLNAGYHF--CGSLISDQWV 59  
 QY 83 TAAHCGNKLMAVGVDDHLL--QGEOLARTRSVYVHKYHSGSPILPRTRDEHML 141  
 DB 60 SAHCKYKSHIQVRLSEVINIDVLENGEAFNSAKVIRHNYSM--IL-----DNDIVLI 111  
 QY 142 KLAAPVPGPRVRAQLPYKCAQPGDQCVAGWGTAAARVYKNGKLTCCSITLLSPDEC 201  
 DB 112 KLSPPAVINARVAITSLPRACAPGTQCIISGMGNTLSSGTYNPBLLOCLDAPVLYTQAC 171

QY 202 EEFYPGVNTNNMTCAG-LDRGDPCCSDSGPVCDETLQGLISKGVYPCGSAQHPAVYT 260  
 DB 172 EASTPQITERNMTCAGLEGGKSDSCGDSGPVACNGLOGLVSMG-YGCAQKQKPGVYT 230  
 QY 261 QICKYMSWINKVIRSN 276  
 DB 231 KVCNVDWIMQETIAAN 246

RESULT 8  
 TRYZ BOVIN STANDARD; PRT; 247 AA.  
 ID TRYZ BOVIN STANDARD; PRT; 247 AA.

AC 029463;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trypsin, anionic precursor (EC 3.4.21.4).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 OC Bovidae; Bovinae; Bos.  
 OK NCBI\_Taxid=9913;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Holstein-Friesian; TISSUE=Pancreas;  
 RX MEDLINE=91065383; PubMed=1701147;  
 RA le Huereu I., Wicker C., Guillebeau P., Toulliec R., Pilgerver A.;  
 RT "Isolation and nucleotide sequence of cDNA clone for bovine  
 pancreatic anionic trypsinogen. Structural identity within the  
 trypsin family.";  
 RL Eur. J. Biochem. 193:767-773(1990).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO TRYPSIN FAMILY 1.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; X54703; CAA38513.1; -  
 DR PIR; S13813; S13813.  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01288;  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydroxylase; Serine protease; Digestion; Pancreas; Zymogen; Signal.  
 FT SIGNAL 1 15  
 FT PROPEP 16 23  
 FT CHAIN 24 247  
 FT ACT\_SITE 63 63  
 FT ACT\_SITE 107 107  
 FT ACT\_SITE 200 200  
 FT DISULFID 30 160  
 FT DISULFID 48 64  
 FT DISULFID 132 233  
 FT DISULFID 139 206  
 FT DISULFID 171 185  
 FT DISULFID 196 220  
 FT SITE 194 194  
 FT SEQUENCE 247 AA; 26289 MM; 50A070495A731DB CEC64;

Query Match 33.8%; Score 505.5; DB 1; Length 247;  
 Best Local Similarity 40.9%; Pred. No. 1,le-38;  
 Matches 105; Conservative 42; Mismatches 95; Indels 15; Gaps 6;

QY 22 PLMAQLMAAABALLPONDRLDPEAYGAPCARSGPQWVSLFNLSPHCAGVLVDQSMV 81  
 DB 3 PLILAFGAFAVAPSSDDDKI---VSGYITAEHNSVPIQVSLNACIYH-CGSLINDQMV 58  
 QY 82 LTAHCGKPKMARVGDHLLLO-GEQLRTTSSVHPKHYOGSGPILPRTEHDLML 140  
 DB 59 VSAHACYGHIQVRLGEYINIDVLEGGQFIDASKIRIRPKSSM-----TLNDIDL 110  
 QY 141 LKLARPVPRPRVALQPRCAQBDQCQVAGNTPARAKYKINKLTCSTITLSPE 200  
 DB 111 IKLSPAVINARVSTLLPSACASAGTECLISGMNLTSSGVNPDLLQCLVAPILSHAD 170  
 QY 201 CEVFPGVNTNNMTCAG-LDRGDPCCSDSGPVCDETLQGLISKGVYPCGSAQHPAVY 259  
 DB 171 CEASYPGQITNNMTCAGLEGGKSDSCGDSGPVACNGLOGLVSMG-YGCAQKQKPGVY 229  
 QY 260 TQICKYMSWINKVIRSN 276  
 DB 230 TKVCNVDWIMQETIAAN 246

## RESULT 9

KKB HUMAN STANDARD; PRT; 250 AA.  
 ID KKB HUMAN STANDARD; PRT; 250 AA.  
 AC Q9UBX7; O75837; Q9NS65;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Kallikrein 11 precursor (EC 3.4.21.-) (Hippocastin) (Trypsin-like  
 protease).  
 GN K1K1 OR PRSS20 OR TISP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OK NCBI\_Taxid=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Hippocampus;  
 RX MEDLINE=98438738; PubMed=9765601;  
 RA Yoshida S., Taniguchi M., Suemoto T., Oka T., He X.P., Shiozaka S.;  
 RT "cDNA cloning and expression of a novel serine protease, TISP.";  
 RL Biochem. Biophys. Acta 1399:225-228(1998).  
 RN (2)  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE=Hippocampus, and Prostate;  
 RX MEDLINE=20329229; PubMed=10872828;  
 RA Mitsui S., Yamada T., Okui A., Koninami K., Uemura H., Yamaguchi N.;  
 RT "A novel isoform of a kallikrein-like protease, TISP/hippocastin,  
 (PRSS20), is expressed in the human brain and prostate.";  
 RL Biochem. Biophys. Res. Commun. 272:205-211(2000).  
 RN (3)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=2030117; PubMed=10662548;  
 RA Yousef G.M., Scoriias A., Diamandis E.P.;  
 RT "Genomic organization, mapping, tissue expression, and hormonal  
 regulation of trypsin-like serine protease (TISP PRSS20), a new  
 member of the human kallikrein gene family.";  
 RL Genomics 63:88-96(2000).  
 RN (4)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=20510030; PubMed=11054574;  
 RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,  
 RA Moss F., Paepel B., Wang K.;  
 RT "Sequencing and expression analysis of the serine protease gene  
 cluster located in chromosome 19q13 region.";  
 RL Gene 257:119-130(2000).  
 RN (5)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Immerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,  
 RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stiliwagen S.,  
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Gaines J.,  
 RA Dangnan L., Exler A., Christensen M., Georgescu A., Avila J., Liu S.,



Genomics 65:184-194(2000).

SEQUENCE FROM N.A.

MEDLINE=20510030; PubMed=11054574;

Gan J., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,

Mose P., Paepke B., Wang K.,

"Sequencing and expression analysis of the serine protease gene

cluster located in chromosome 19q13 region."

Gene 257:119-130(2000).

SEQUENCE FROM N.A.

Lamerdin J.E., McCreedy P.M., Skowronski E., Vismathan V.,

Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilaegen S.,

Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Ganes J.,

Dangnan L., Trankheim M., Christensen M., Georgescu A., Avila J., Liu S.,

Duarte S., Lucas S., Bruce R., Amico-Keller G., Coefield J.,

Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,

Olsen A.S., Carrano A.V.;

"Sequence analysis of chromosome 19q13.4."

Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.

-1- SUBCELLULAR LOCATION: Secreted (probable).

-1- TISSUE SPECIFICITY: SKIN, THYMUS, TRACHEA, CEREBELLUM AND SPINAL

CORD.

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. KALLIKREIN SUBFAMILY.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>

or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

-----

EMBL: AF135026; AAD26427.2; -

EMBL: AF243527; AAG33362.1; -

EMBL: AC011473; AAG32355.1; -

HSPB; P00763; IDPO.

MEROFS; S01.307; -

Genew: HGNC:6370; KLK9.

MM: 605504; -

GO: GO:0005576; C:extracellular; NAS.

GO: GO:0006508; P:proteolysis and peptidolysis; NAS.

InterPro: IPR001314; Chymotrypsin.

InterPro: IPR001254; Ser protease\_Try.

Pfam: PF00089; trypsin\_1.

PRINTS: PR00722; CHYMOTRYPSIN.

SMART: SM00020; TRYP\_SPC; 1.

PROSITE: PS50240; TRYP\_SIN\_DOM; 1.

PROSITE: PS00134; TRYP\_SIN\_HIS; 1.

PROSITE: PS00135; TRYP\_SIN\_SER; 1.

Hydrolase; Serine protease; Glycoprotein; signal.

FT SIGNAL 1 15

FT CHAIN 1 250

FT ACT\_SITE 63 250

FT ACT\_SITE 111 111

FT ACT\_SITE 204 204

FT DISULFID 29 164

FT DISULFID 48 64

FT DISULFID 136 238

FT DISULFID 143 210

FT DISULFID 175 189

FT DISULFID 200 225

FT CARBOHYD 131 131

FT CARBOHYD 166 166

FT CARBOHYD 211 211

SEQUENCE 250 AA; 27512 MW; F2785245B063E8B8 CRC64;

Query Match 33.6%; Score 502; DB 1; Length 250;

Best Local Similarity 39.7%; Pred. No. 2.3e-38;

Matches 104; Conservative 39; Mismatches 99; Indels 20; Gaps 6;

17 LAKLPLMAQMAEAALPQNDRLDPEAIGA-PCARGSPQVQVSLFNGSLFHCAVYL 75

Db 6 LKALLSLAGHNA-----DTR-----AIGAEBCRPNSQPMQAGLFLTRLFCGATL 52

Qy 76 VDSWVLTAAHCNKPLMARVGDHLLLOG-FQLKRTTSVHPKXHGSSGPIPRRTD 134

Db 53 ISRMMLTAAHCNKPLMARVGDHLLLOG-FQLKRTTSVHPKXHGSSGPIPRRTD 108

Qy 135 EHDMLKLARPVVPVPRVATLQPYRCAPGQCCVAGGTAAARVKNKGLTCSIT 194

Db 109 NDDIMLRPLPRQARLSPAVQPLNLSTQVSPGHCCLISGMAVSPKALFPVTLQCANIS 168

Qy 195 ILSPKCEVFPVGVVNTNMICAGL-DRGDDPCQSDSGPLVCDTELQILSWGVPCGSA 253

Db 169 ILNKLCHAVYPPHIDSMCLAGLMEGRGSCGDSGGLVNGTLAGVVGAGAPCSR 228

Qy 254 QHNAVYQICKWSMINKYIRS 275

Db 229 RRPVAVTSVCHYDWTQETMEN 250

RESULT 11

KLKE\_HUMAN STANDARD; PRT; 251 AA.

AC Q9P0G3;

DT 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Kallikrein 14 precursor (EC 3.4.21.-) (Kallikrein-like protein 6)

DE (KLK-16).

GN KLK14 OR KLK16

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

NP [1]

RP SEQUENCE FROM N.A.

RA Yousif G.M., Diamandis E.P.;

RT "Molecular characterization, mapping, and tissue expression of KLK16,

RT a hormonally regulated kallikrein-like gene."

RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.

RM [2]

RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.

RA MEDLINE=21250997; PubMed=11352573;

RA Hooper J.D., Bul J.T., Rae F.K., Harvey T.J., Myers S.A.,

RA Ashworth L.K., Clements J.A.;

RT "Identification and characterization of KLK14, a novel kallikrein

RT serine protease gene located on human chromosome 19q13.4 and expressed

RT in prostate and skeletal muscle."

RL Genomics 73:117-122(2001).

RM [3]

RP SEQUENCE FROM N.A.

RA Lamerdin J.E., McCreedy P.M., Skowronski E., Vismathan V.,

RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilaegen S.,

RA Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Ganes J.,

RA Dangnan L., Trankheim M., Christensen M., Georgescu A., Avila J., Liu S.,

RA Andreise T., Trankheim M., Attix C., Amico-Keller G., Coefield J.,

RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,

RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,

RA Olsen A.S., Carrano A.V.;

RT "Sequence analysis of chromosome 19q13.4."

RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.

RM [4]

RP TISSUE SPECIFICITY.

RA MEDLINE=20545474; PubMed=10969073;

RA Harvey T.J., Hooper J.D., Myers S.A., Stephenson S.A., Ashworth L.K.,

RA Clements J.A.;

RT "Tissue-specific expression patterns and fine mapping of the human

RT kallikrein (KLK) locus on proximal 19q13.4."

RL J. Biol. Chem. 275:37397-37406(2000).

CC -1- SUBCELLULAR LOCATION: Secreted (probable).

CC -1- TISSUE SPECIFICITY: HIGH EXPRESSION IN LIVER, PANCREAS, FETAL SPLEEN,

CC PROSTATE AND SKELETAL MUSCLE.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. KALLIKREIN SUBFAMILY.  
 CC -----  
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF161221; AAD50773.2; -  
 DR EMBL; AF283669; AAK48523.1; -  
 DR EMBL; AF283670; AAK48524.1; -  
 DR EMBL; AC011473; AAG23260.1; -  
 DR HSP; P00763; IDPO.  
 DR MEROPS; S01.029; -  
 DR Genew; HGNC:6362; KLK14.  
 DR MIM; 606135; -  
 DR GO; GO:0005576; C.electraceutical; NAS.  
 DR GO; GO:0006508; P.proteolysis and peptidolysis; NAS.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser. protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR KMW; KMW:KMW:0135; Serine protease; Signal; Zymogen.  
 FT SIGNAL; 1 18  
 FT PROPEP; 19 24  
 FT CHAIN; 25 251  
 FT ACT\_SITE; 67 67  
 FT ACT\_SITE; 111 111  
 FT ACT\_SITE; 204 204  
 FT DISULFID; 31 164  
 FT DISULFID; 52 68  
 FT DISULFID; 143 210  
 FT DISULFID; 175 189  
 FT DISULFID; 200 225  
 SQ SEQUENCE 251 AA; 27452 MW; 9087953BAF7ED25 CRC64;  
 Query Match 33.4%; Score 500; DB 1; Length 251;  
 Best Local Similarity 42.2%; Pred. No. 3.5e-38;  
 Matches 97; Conservative 36; Mismatches 85; Indels 12; Gaps 4;  
 QY 49 GAPCARGSQPQVSLFNG--LSFHCAGVLVDQSWLTAACNCKPLMARVGDHLLILQ 106  
 DB 28 GHTCTRSSQPMQALLAGPRRRFLCGALLSGQWITAHGCRPIQLVAGNLRMEA 87  
 QY 107 -EOLARTTSVVAHKYHQSGLPRRTDEHMLMLKLARPVVPGRRVALQLPYRCAP 165  
 DB 88 TQGVLRVVAQVTHPENYS-----RTHDNDMLLOLQOPARIGRAVRPIEVTOACASP 139  
 QY 166 GDQCVAGGTTAARVKNKGLTCSITLISPEKCEVYPPVANNMTCGLDR-GDDP 224  
 DB 140 GTSCHVSGMGTISSTIARPAALQGVNINISPDVQCAQAFPTITPGWAGVPGGKDS 199  
 QY 225 CGSDSGPIVCDDELQGLISGCVTPCGSAQHPAVYTOICKWMSWINKYR 274  
 DB 200 CGSDSGPIVCRGOLGLVSWGMRICALPGYGVYTNLCIKRYSWIEEMR 249  
 RESULT 12  
 NRPN RAT STANDARD; PRT; 260 AA.  
 AC 088780;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neutropsin precursor (BC 3.4.21.-) (NP) (Kallikrein 8) (Brain serine  
 DE protease 1).  
 GN KLK8 OR PRSS19 OR NRPN OR BSP1.

OS Rattus norvegicus (Rat).  
 CC Burkarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC NCB1\_TaxID=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Fischer; Tissue=Brain;  
 RX MEDLINE=98389725; PubMed=972524;  
 RT Davies B.J., Pickard B.S., Steel M., Morris R.G.M., Latche R.;  
 RL J. Biol. Chem. 273:23004-23011(1998).  
 CC -1- FUNCTION: SUGGESTED TO BE INVOLVED IN KINDLING EPILEPTOGENESIS AND  
 CC HIPPOCAMPAL PLASTICITY. HAS A STRONG PROTEOLYTIC ACTIVITY AGAINST  
 CC FIBRONECTIN (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-, Lys-.  
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -1- TISSUE SPECIFICITY: RESTRICTED TO HIPPOCAMPS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. KALLIKREIN SUBFAMILY.  
 CC -----  
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ006641; CA06643.1; -  
 DR HSP; Q01955; INPM.  
 DR MEROPS; S01.244; -  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser. protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR KMW; KMW:KMW:0135; Serine protease; Glycoprotein; Zymogen; Signal.  
 FT SIGNAL; 1 28  
 FT PROPEP; 29 32  
 FT CHAIN; 33 260  
 FT ACT\_SITE; 73 73  
 FT ACT\_SITE; 120 120  
 FT ACT\_SITE; 212 212  
 FT DISULFID; 39 173  
 FT DISULFID; 58 74  
 FT DISULFID; 145 246  
 FT DISULFID; 152 218  
 FT DISULFID; 184 198  
 FT DISULFID; 208 233  
 FT CARBOHYD; 110 110  
 SQ SEQUENCE 260 AA; 28510 MW; 58DF4F062A0B7F5 CRC64;  
 Query Match 33.4%; Score 499.5; DB 1; Length 260;  
 Best Local Similarity 39.7%; Pred. No. 4e-38;  
 Matches 100; Conservative 38; Mismatches 105; Indels 9; Gaps 4;  
 QY 23 LMAQLAAAPALLPQNDTRLDPEAYGAPCARSGPQVSLFNGLSFHCAGVLVDQSWL 82  
 DB 13 ILFLMGAMVAGLTRAQSKT---LKGQCEKPHSQPMQALFQGERLVCGGVLDGRVYL 69  
 QY 83 TAAHQGNKPLMARVGDHLLIL--QGEQLARTTSVVAHKYHQSGLPLPRRTDEHML 141  
 DB 70 TAAHCKDKYSVRLGDSLSQRRDEPEOIRVANSIQPCFN--SNP---EDHSHDML 124  
 QY 142 KLAAPVVGRRVVALQLPYRCAPGDCQVAGMGTTAARVKNKGLTCSITLISPEK 201  
 DB 125 RLQNSAMLGKVRKRIELANLCPKVGKCIISGNGTVLSPENPNTLNCABEVKYSQNK 184  
 QY 202 EVFPGVYTNMTCAGLDRQDPQSGSGGPIVCDDELQGLISGCVTPCGSAQHPAVYTO 261  
 DB 185 BRAPGKITGKVCAGSSNGADTCCGSDSGPIVCRGOLGLVSWGMRICALPGYGVYTNLCIKR 244

QY 262 ICKYMSWINKY 273  
DB 245 ICRYTNWIKKTM 256

## RESULT 13

ID KLIK\_HUMAN STANDARD; PRT; 277 AA.

AC Q9UKR3; Q9Y433;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Kallikrein 13 precursor (EC 3.4.21.-) (Kallikrein-like protein 4)  
DE (KLIK-14).

GN KLIK3 OR KLIK4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;

## SEQUENCE FROM N.A.

RA MEDLINE=20229789; PubMed=10766816;

RA Yousef G.M., Chang A., Diamandis E.P.;

RT "Identification and characterization of KLIK-14, a new kallikrein-like gene that appears to be down-regulated in breast cancer tissues.";

RL J. Biol. Chem. 275:11891-11898(2000).

RN [2]

## SEQUENCE FROM N.A.

RA Lamedin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,

RA Burkhart-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,

RA Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Ganes J.,

RA Dangnan T., Frankheim M., Christensen M., Georgescu A., Avila J., Liu S.,

RA Duarte S., Lucas S., Bruce R., Altix C., Amico-Keller G., Coefield J.,

RA Acellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,

RA Olsen A.S., Carraro A.V.;

RT "Sequence analysis of chromosome 19q13.4";

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

## SEQUENCE OF 1-180 FROM N.A.

RA TISSUE=Dermis;

RA Ansoorge W., Winkner U., Mewes H.-W., Gassenhuber J., Wiemann S.;

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Secreted (Probable).

CC -1- TISSUE SPECIFICITY: EXPRESSED IN PROSTATE, BREAST, TESTIS AND

SALIVARY GLAND.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. KALLIKREIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL; AF135024; AAD26425.2; -

DR EMBL; AC011473; AAC23259.1; -

DR EMBL; AL050220; CAB43320.1; ALT\_INT.

DR HSSP; P00763; IDPO.

DR MEROPS; S01.306; -

DR GeneW; HGNC:6361; KLIK3.

DR MIM; 605505; -

DR GO; GO:0005576; C:extracellular; NAS.

DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.

DR InterPro; IPR001254; Ser\_protease\_Try.

DR Pfam; PF00089; trypsin; 1.

DR SMART; SM00020; TRYP\_SPC; 1.

DR PROSITE; PS00240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

FT SIGNAL 1 16  
FT CHAIN 17 277  
FT ACT\_SITE 76 76  
FT ACT\_SITE 124 124  
FT ACT\_SITE 218 218  
FT DISULFID 42 178  
FT DISULFID 61 77  
FT DISULFID 157 224  
FT DISULFID 157 224  
FT DISULFID 189 203  
FT DISULFID 214 239  
FT CARBOHYD 30 30  
FT CARBOHYD 225 225  
FT CONFLICT 170 180  
SQ SEQUENCE 277 AA; 30570 MW; BAA9A9EDCEB50542 CRC64;

Query Match 33.3%; Score 497.5; DB 1; Length 277;  
Best Local Similarity 43.2%; Pred. No. 6; Se-38;  
Matches 99; Conservative 38; Mismatches 85; Indels 7; Gaps 4;

QY 49 GAPCARGSQPMQVSLFNGSLFACGLVLDQSWVLTAAHGNKFLMARVGDHLLLLQ-GE 107  
DB 39 GYTCFPHSQPMQVSLFNGSLFACGLVLDQSWVLTAAHGNKFLMARVGDHLLLLQ-GE 98

QY 108 QKRTTSVHPKHYQSGPILPRRTDEHMLKLARVVGPRVRLQLPY-RCQAPG 166  
DB 99 QREVVHSHIPHEVRRSPFTHL----NHDDIMLELQSPVQLTGYLTPLSHNNRLRFG 154

QY 167 DCCOVAGMGTARAVKYNKGLTCSITLSPKEVEFPRGVVNTMTCAGL-DGQDPC 225  
DB 155 TTCRSGMGTITSPVNTPKTLQCANIQRSDEEROYVPGKITNMCAQKGGKQSC 214

QY 226 QSDSGPLVCDETLQGLISWGVYPCGSAQHPAVYTIQCKYMSWINKVR 274  
DB 215 ESDSGPLVCNNTLGLVSWGDFPCGPDREGVYRVRVYIMRETR 263

RESULT 14  
TRYL\_BOVIN  
ID TRYL\_BOVIN STANDARD; PRT; 243 AA.  
AC P00760;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Trypsinogen, cationic precursor (EC 3.4.21.4) (Beta-trypsin)  
DE (Fragment).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RA TISSUE=Pancereas;  
RA Okajima T., Maniwa M., Nagao S., Fujikawa H., Goto S.;

RA Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.

RN [2]  
RA MEDLINE=67168848; PubMed=5967094;  
RA Mikes O., Holeysovsky V., Tomasek V., Sorm F.;

RT "Covalent structure of bovine trypsinogen. The position of the remaining amides";

RT Biochem. Biophys. Res. Commun. 24:346-352(1966).

RN [3]  
RA MEDLINE=72035053; PubMed=4399051;  
RA Hartley B.S.;

RT "Homologies in serine proteinases";

RT Phillos. Trans. R. Soc. Lond., B, Biol. Sci. 257:77-87(1970).

RN [4]  
RA MEDLINE=75146445; PubMed=1092332;  
RA Tlanti K., Ericsson L.H., Neurath H., Walsh K.A.;

RT "Amino acid sequence of dogfish trypsin.";







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 08:58:10 / Search time 27 Seconds

(without alignments)  
983.058 Million cell updates/sec

Title: US-10-021-368-1

Sequence: 1496  
1 MEAPHILHSASASGAPALAKL.....AVYIQCKMSWINKVIRSN 276

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96166682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the local score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	549	36.7	253	2 A53968	serine proteinase
2	513	34.3	260	2 I56559	neuropilin - mouse
3	506	33.8	247	1 TRDG	trypsin (EC 3.4.21
4	505.5	33.8	247	2 S13813	trypsin (EC 3.4.21
5	494	33.0	246	2 B25628	trypsin (EC 3.4.21
6	491	32.8	229	1 TRB0TR	trypsin (EC 3.4.21
7	488.5	32.7	246	1 TRRT2	trypsin (EC 3.4.21
8	486.5	32.5	247	2 S05494	trypsin (EC 3.4.21
9	483	32.3	231	1 TRP0TR	trypsin (EC 3.4.21
10	482	32.2	246	1 TRRT1	trypsin (EC 3.4.21
11	477.5	31.9	247	2 A27547	trypsin (EC 3.4.21
12	473.5	31.7	242	2 S49489	trypsin (EC 3.4.21
13	470	31.4	247	1 B25852	trypsin (EC 3.4.21
14	469.5	31.4	248	2 S35066	trypsin (EC 3.4.21
15	468.5	31.3	238	2 S31779	trypsin (EC 3.4.21
16	468.5	31.3	304	2 S33496	trypsin (EC 3.4.21
17	467	31.2	229	1 TRDS	trypsin (EC 3.4.21
18	464.5	31.0	259	2 I38363	trypsin (EC 3.4.21
19	460	30.7	246	1 TRDG	trypsin (EC 3.4.21
20	458.5	30.6	247	1 A25852	trypsin (EC 3.4.21
21	457.5	30.6	261	2 A31336	trypsin (EC 3.4.21
22	457	30.5	243	2 A35871	trypsin (EC 3.4.21
23	453	30.3	248	2 S55067	trypsin (EC 3.4.21
24	450.5	30.1	231	2 S31778	trypsin (EC 3.4.21
25	450	30.1	260	2 A37938	trypsin (EC 3.4.21
26	445.5	29.8	242	2 S31775	trypsin (EC 3.4.21
27	445.5	29.8	242	2 S31776	trypsin (EC 3.4.21
28	443	29.6	240	2 S39047	trypsin (EC 3.4.21
29	439.5	29.4	261	2 S45503	trypsin (EC 3.4.21

30	438	29.3	247	2 S12764	trypsin (EC 3.4.21
31	436	29.1	241	2 S39048	trypsin (EC 3.4.21
32	436	29.1	259	2 B31336	cisue kallikrein
33	434.5	29.0	256	1 MGMSA	7S nerve growth fa
34	431.5	28.8	244	2 A44284	cisue kallikrein
35	429	28.7	261	1 K0MS1	cisue kallikrein
36	427.5	28.6	261	2 A34079	cisue kallikrein
37	426.5	28.5	259	2 D23863	cisue kallikrein
38	425.5	28.4	239	2 A27207	cisue kallikrein
39	425	28.4	232	1 K0RG	cisue kallikrein
40	425	28.4	261	2 A29745	cisue kallikrein
41	424	28.3	246	2 J01471	trypsin (EC 3.4.21
42	424	28.3	259	1 K0RTN	comin (EC 3.4.21.-
43	423	28.3	246	2 J01472	trypsin (EC 3.4.21
44	423	28.3	265	1 K0RTP	cisue kallikrein
45	422	28.2	261	2 A25606	cisue kallikrein

#### ALIGNMENTS

##### RESULT 1

A53968

serine proteinase SCCE precursor - human

C:Species: Homo sapiens (man)

C>Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #ext\_change 22-Jun-1999

C/Accession: A53968

R/Hansson, L.; Stromqvist, M.; Baeckman, A.; Wallbrant, P.; Carlstein, A.; Egelund, J. Biol. Chem. 269, 19420-19426, 1994

A:Title: Cloning, expression, and characterization of seratum corneum chymotryptic enz

A:Reference number: A53968, PMID:94308225, PMID:8034709

A/Accession: A53968

A>Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-253 <HAN>

A/Cross-references: GB:133404; MID:9521214; PIDN:AAC37551.1; PID:9532504

C/Genetics:

A/Gene: GDB:PRSS6; SCCE

A/Cross-references: GDB:377730

A/Map position: 7q35-7q35

C/Superfamily: trypsin; trypsin homology

F/30-245/Domain: trypsin homology <TRY>

Query Match 36.7%; Score 549; DB 2; Length 253;  
Best local similarity 40.9%; Pred. No. 1.7e+40;  
Matches 108; Conservative 45; Mismatches 97; Indels 14; Gaps 4;

QY	14	ARALAKLPLMAQMAADALIPQNDRLDPBAYGAPCARGSQPMQVSLFNGLSFHCG	73
DB	2	ARSLILPLQTLISLALSTRAGEBAQGDKITD---GAPCARGSHPQVALLSGNQLHCCG	57
QY	74	VLVDQSVLTPAAGCNKPLMARVGDHLLLOEOLRRTTRSVHPKYGSGPLIPRT	133
DB	58	VLNERNVLTFAACRKNMETYVHIGSPTLGDRAQRV-KASKSPRHPGV-----STGT	108
QY	134	DEHDMILKARPVVGPFRALQLPYRCAQPDQCVAGMGTTAARVYKNGLTCSGI	193
DB	109	HVMDIMLVKINSQARLSWYKVRLEPCEPFTCTVSGMGTTSPDVTFPSDLMCVDV	168
QY	194	TIISPKCEVFYFGVNTNMICAGL--DRGDDPQSGSGPLVCDELQGLISNGVYPCGS	252
DB	169	KLSFQDCTKVYTDLLSNMLCAGIDPSKKNACNGSGGLVCRGLTQGLVSWGTFCQ	228
QY	253	AQHPAVYQICKYMSWINKVIRSN	276
DB	229	PNDGVTYQVCKRFTKWINDTMKGH	252

##### RESULT 2

I56559

neuropilin - mouse  
C:Species: Mus musculus (house mouse)

C>Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 21-Jul-2000  
 C/Accession: I56559  
 R/Chem, Z.L.; Yoshida, S.; Kato, K.; Momota, Y.; Suzuki, J.; Tanaka, T.; Ito, J.; Nishih  
 J. Neurosci. 15, 5088-5097, 1995  
 A>Title: Expression and activity-dependent changes of a novel limbic-serine protease gen  
 A/Reference number: I56559; MUID:9534817; PMID:7623137  
 A/Accession: I56559  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-260 <RES>  
 A/Cross-references: GB:D30785; NID:G1648847; PIDN:BAA06451.1; PID:G1020921  
 C/Superfamily: trypsin; trypsin homology  
 F:33-252/Domain: trypsin homology <TR>

Query Match 34.8%; Score 513; DB 2; Length 260;

Best Local Similarity 41.6%; Pred. No. 2,4e-37;  
 Matches 106; Conservative 32; Mismatches 105; Indels 12; Gaps 5;

QY 20 LLLPILMAQLMAAEALLPONDTRLDPEAYGAPCARSGQPMQVSLFNGLSFHCAGVLVDQS 79  
 13 ILLILFPGANA--GLTRAQSKILE---GRECIPIHQPMQALFQGBRLICGGVLVGR 66  
 QY 80 WVLTAHCGNKPIMARVGDHLLIL-QGEQLRRTTSVAPKTHQSGGPIPRTEHDL 138  
 67 WVLTAHCGNKPIMARVGDHLLIL-QGEQLRRTTSVAPKTHQSGGPIPRTEHDL 121  
 QY 139 MLKLARPVVGPVRVALOLPYRCAGPQDQCOVAGMGTATARRVKYKNGLTGSSITILSP 198  
 122 MLILQNSANLGDVKVQVQLANLCPKVGQKCIISGKTIVSPENPNTLNCHEVATISQ 181  
 QY 199 KECEVFPVGVVNTNMICAGLDRGDPCQSDSGGLVDETLQGLISWGVPCGSAQHPAV 258  
 182 NKCRAPVFGKITBEMVACGSSNGADTCQSDSGGLVDCMLQGITSGSDPCGKPEKPV 241  
 QY 259 YTOICKYMSWINKYI 273  
 242 YTKICRYTWIKTKM 256

## RESULT 3

TRDG  
 trypsin (BC 3.4.21.4) precursor, anionic - dog  
 N/Alternate names: cationic trypsinogen  
 C/Species: Canis lupus familiaris (dog)  
 C/Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 18-Jun-1999  
 C/Accession: A26273  
 R/Pinsky, S.D.; LaForge, K.S.; Scheele, G.  
 Mol. Cell. Biol. 5, 2669-2676, 1985  
 A>Title: Differential regulation of trypsinogen mRNA translation: full-length mRNA sequ  
 A/Reference number: A26273; MUID:86284626; PMID:38411794  
 A/Accession: A26273  
 A/Molecule type: mRNA  
 A/Residues: 1-247 <PIN>  
 A/Cross-references: GB:M11589; NID:G164094; PIDN:AAA30899.1; PID:G164095  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:16-23/Domain: activation peptide #status predicted <APR>  
 F:24-247/Product: trypsin, anionic #status predicted <ENZ>  
 F:24-239/Domain: trypsin homology <TRY>  
 F:30-160,48-64,132-239,139-206,171-185/Disulfide bonds: #status predicted  
 F:63,107,200/Active site: His, Asp, Ser #status predicted  
 F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 33.8%; Score 506; DB 1; Length 247;  
 Best Local Similarity 41.8%; Pred. No. 9.3e-37;  
 Matches 107; Conservative 41; Mismatches 92; Indels 16; Gaps 8;

QY 23 LLMQLMAAEALLPONDTRLDPEAYGAPCARSGQPMQVSLFNGLSFHCAGVLVDQS 82  
 5 LILARLGAAYAT--PTDD--DKIVGVTCENSVPVQVSLNKGYNH-CGGSLSIDQWV 59  
 QY 83 TAAHCGNKPIMARVGDHLLILQ-GEQLRRTTSVAPKTHQSGGPIPRTEHDLML 141

DB 60 SAACVYSRIQVRAGEYNIDLEGNEQFINSAVIRPWNNSV--IL-----DNDIMLI 111  
 QY 142 KLARPVVGPVRVALOLPYRCAGPQDQCOVAGMGTATARRVKYKNGLTGSSITILSPKEC 201  
 112 KLSPAVLNRVATISLPRACAPRGYGLISWGVNTLSSTNYPPELLQCLDARILTQAQC 171  
 QY 202 EVFYPGVVNTNMICAG-LDRGDDPCQSDSGGLVDETLQGLISWGVPCGSAQHPAVYT 260  
 172 EASYPGQITENMILCAGLEGGKQSCQSDSGGLVDCMLQGITSGSDPCGKPEKPVYT 230  
 QY 261 QICKYMSWINKYIRSN 276  
 231 KVCNFPVDMIGSTIAN 246

RESULT 4  
 S13813  
 trypsin (BC 3.4.21.4) - bovine

C/Species: Bos primigenius taurus (cattle)  
 C/Date: 02-Dec-1993 #sequence\_revision 03-Aug-1995 #text\_change 22-Jun-1999  
 C/Accession: S13813  
 R/Le Huecrou, I.; Wicker, C.; Guilloteau, P.; Toullec, R.; Puigserver, A.  
 Eur. J. Biochem. 193, 767-773, 1990  
 A>Title: Isolation and nucleotide sequence of cDNA clone for bovine pancreatic anionic  
 A/Reference number: S13813; MUID:91065383; PMID:1701147  
 A/Accession: S13813  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-247 <HUB>  
 A/Cross-references: EMBL:X54703; NID:9829; PIDN:CAA38513.1; PID:9830  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; protein digestion; serine proteinase  
 F:24-239/Domain: trypsin homology <TRY>  
 F:63,107,200/Active site: His, Asp, Ser #status predicted

Query Match 33.8%; Score 505.5; DB 2; Length 247;  
 Best Local Similarity 40.9%; Pred. No. 1e-36;  
 Matches 105; Conservative 42; Mismatches 95; Indels 15; Gaps 6;

QY 22 PLMAQLMAAEALLPONDTRLDPEAYGAPCARSGQPMQVSLFNGLSFHCAGVLVDQSV 81  
 3 PLILAPFAGAVAPFSDDDKXI---VGGYCAENSVYQVSLNKGYNH-CGGSILNDQWV 58  
 QY 82 LTAHCGNKPIMARVGDHLLILQ-GEQLRRTTSVAPKTHQSGGPIPRTEHDLML 140  
 59 VSAHCVYTHIQVRIGEVNIDLEGGEQFIDASKIRHPKYSW-----TLNDIMLI 110  
 QY 141 LKLARPVVGPVRVALOLPYRCAGPQDQCOVAGMGTATARRVKYKNGLTGSSITILSPKE 200  
 111 IKLSTPAINARVSTILPLPSACASAGTECLISGNTLSGVNYPDLQCLVAPLISHAD 170  
 QY 201 CEVFPVGVVNTNMICAG-LDRGDDPCQSDSGGLVDETLQGLISWGVPCGSAQHPAVYT 259  
 171 CEASYPGQITENMILCAGLEGGKQSCQSDSGGLVDCMLQGITSGSDPCGKPEKPVYT 229  
 QY 260 QICKYMSWINKYIRSN 276  
 230 TKVCNFPVDMIGSTIAN 246

## RESULT 5

B25528  
 trypsin (BC 3.4.21.4) precursor - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 22-Jun-1999  
 C/Accession: B25528  
 R/Stevenson, B.J.; Hagenbuchle, O.; Wellauer, P.K.  
 Nucleic Acids Res. 14, 8307-8330, 1986  
 A>Title: Sequence organisation and transcriptional regulation of the mouse elastase I  
 A/Reference number: B25528; MUID:87066713; PMID:3641189  
 A/Accession: B25528  
 A/Molecule type: mRNA

A:Residues: 1-246 <STB>  
 A:Cross-references: GB:X04574; NID:954918; PIDN:CAA28243.1; PID:954919  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: calcium binding; hydrolase; protein digestion; serine proteinase  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-246/Product: trypsin #status predicted <MAT>  
 F:24-239/Domain: trypsin homology <TRY>  
 F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted  
 F:63,107,200/Active site: His, Asp, Ser #status predicted  
 F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 33.0%; Score 494; DB 2; Length 246;  
 Best Local Similarity 41.0%; Pred. No. 1e-35;

Matches 105; Conservative 40; Mismatches 95; Indels 16; Gaps 7;  
 QY 23 LLMALWMAZAAALPONDRLDPEAYGAPCARSGQPMQVSLFNGLSFHCAGVLDQSWYL 82  
 DB 5 LILALVGAAYV--FPVDD--DKIVGYTCRESSVPYQVSLNAGYHF--CGGSLINQWV 59  
 QY 83 TAAAGCKKPELMARVGDHLLILQG--EQLRRTTSVVPKXHQSGPILPRRTDHDML 141  
 DB 60 SAACIKYKRIQVRIGSEHNINVLGEINQFVDSAKIRHPNNSW-----TLDNDIMDI 111  
 QY 142 KLARPVVGPFRVALQIPYRCAQPGDQVAGMGTARVKKYKNGITCSITILSPKEC 201  
 DB 112 KLASPVTLNARVAVPSPSCAPAGTQCLISGMKNTLSNKNPDLQCYDAVLPQADG 171  
 QY 202 EEPYRGVVTNNMTCAG--LDRGQDPCQSDSGGPLVCDETLGILISWGVYPCGSAQHPAVYT 260  
 DB 172 EASYRGDITNNMTCVGLFEGKSCQCGDSGPPVCCNDELGIIVSWG--YCGAQADPAGVYT 230  
 QY 261 QICKYMSWINKVIRSN 276  
 DB 231 KVCNVVDMIQITLADN 246

## RESULT 6

TRBOTR

trypsin (EC 3.4.21.4) precursor - bovine

N:Contains: trypsinogen

C:Species: Bos primigenius taurus (cattle)

C:Date: 24-Apr-1994 #sequence\_revision 28-Feb-1986 #text\_change 18-Jul-1997

C:Accession: A90164; A00946; S08774

R:McKee, O.; Holeyovsky, V.; Tomasek, V.; Sorm, F.

Biochem. Biophys. Res. Commun. 24, 346-352, 1966

A:Title: Covalent structure of bovine trypsinogen. The position of the remaining amides.

A:Reference number: A90164; MUID:67168848; PMID:5967094

A:Accession: A90164

A:Molecule type: protein

A:Residues: 1-27, 'Q', 59-67, 'Q', 69-150, 'N', 152-176, 'N', 178-229 &lt;NIK&gt;

R:Hartley, B.S.

Philos. Trans. R. Soc. Lond. B257, 77-87, 1970

A:Reference number: A93755

A:Contents: annotation; revisions

R:Titani, K.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.

Biochemistry 14, 1358-1366, 1975

A:Title: Amino acid sequence of dogfish trypsin.

A:Reference number: A00950; MUID:75146445; PMID:1092332

A:Contents: annotation; revisions

A&gt;Note: the sequence agrees with that shown

R:Bode, W.; Schwager, P.

J. Mol. Biol. 98, 693-717, 1975

A:Title: The refined crystal structure of bovine beta-trypsin at 1.8 angstrom resolution

A:Reference number: A92954; MUID:76072097; PMID:512

A:Contents: annotation; X-ray crystallography; binding sites for calcium, substrate, and

C:Comment: Autocatalytic cleavage after Lys-6 leads to beta-trypsin by releasing a termi

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen

F:1-229/Product: trypsinogen #status experimental &lt;ZYM&gt;

F:1-222/Domain: activation peptide #status experimental &lt;APR&gt;

F:7-131,132-229/Product: alpha-trypsin #status experimental <MPR>  
 F:6-7/Cleavage site: Lys-116 (enteropeptidase) #status experimental  
 F:13-143,31-47,115-216,122-189,154-168,179-203/Disulfide bonds: #status experimental  
 F:46,90,183/Active site: His, Asp, Ser #status experimental  
 F:58,60,63,68/Binding site: calcium (Glu, Asn, Val, Glu) #status experimental  
 F:131-132/Cleavage site: Lys-Ser (autolytic) #status experimental

Query Match 32.8%; Score 491; DB 1; Length 229;  
 Best Local Similarity 41.7%; Pred. No. 1.8e-35;

Matches 96; Conservative 38; Mismatches 84; Indels 12; Gaps 5;

QY 49 GAPCARSGQPMQVSLFNGLSFHCAGVLDQSWILTAHCGNKPMARVGDHLLILQG-E 107  
 DB 10 GYTCGANVVPYQVSLNAGYHF--CGGSLINQWVSAACIKYKSGIQVRLGEDNINIVEGNE 68  
 QY 108 QLRRTTSVVPKXHQSGPILPRRTDHDMLKLARPVVGPFRVALQIPYRCAQPGD 167  
 DB 69 QFASNSIYHPSINS-----NTLNDIMLKLASASLSNRSVASISLPTSCAAGT 120  
 QY 168 QCVVAGMGTARVKKYKNGITCSITILSPKCEVEYFPGVVTNNMTCAG--LDRGQDPCQ 226  
 DB 121 QCLISGMKNTLSNKNPDLQCYDAVLPQADG 180  
 QY 227 SDGSPVLCDETLGILISWGVYPCGSAQHPAVYTQICKMSWINKVIRSN 276  
 DB 181 GDSGSPVCCSGRLQGIIVSWG--SGCAQKPKRGVYTKVCNVSWIKQITASN 229

## RESULT 7

TRB2

trypsin (EC 3.4.21.4) II precursor - rat

N:Alternate names: trypsinogen II

C:Species: Rattus norvegicus (Norway rat)

C:Date: 05-Apr-1985 #sequence\_revision 30-Sep-1987 #text\_change 18-Jul-1997

C:Accession: A22657; A00949

R:Craik, C.S.; Choo, Q.L.; Swift, G.H.; Quinto, C.; MacDonald, R.J.; Rutter, W.J.

J. Biol. Chem. 259, 14255-14264, 1984

A:Title: Structure of two related rat pancreatic trypsin genes.

A:Reference number: A22657; MUID:85054880; PMID:6094547

A:Accession: A22657

A:Molecule type: DNA

A:Residues: 1-246 &lt;CRA&gt;

R:MacDonald, R.J.; Stary, S.J.; Swift, G.H.

J. Biol. Chem. 257, 9724-9732, 1982

A:Title: Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide sequences

A:Reference number: A00948; MUID:82265624; PMID:6856710

A:Accession: A00949

A:Molecule type: mRNA

A:Residues: 9-246 &lt;MAC&gt;

C:Comment: The trypsin II mRNA is present in much lower quantities than the trypsin I

C:Genetics:

A:introns: 14/1, 67/2

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen

F:1-15/Domain: signal sequence #status predicted &lt;SIG&gt;

F:16-23/Domain: activation peptide #status predicted &lt;APR&gt;

F:24-246/Product: trypsin II #status predicted &lt;ENZ&gt;

F:24-239/Domain: trypsin homology &lt;TRY&gt;

F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted

F:63,107,200/Active site: His, Asp, Ser #status predicted

F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 32.7%; Score 488.5; DB 1; Length 246;  
 Best Local Similarity 39.8%; Pred. No. 3.1e-35;

Matches 105; Conservative 43; Mismatches 95; Indels 21; Gaps 8;

QY 15 PALAKLPLMAQMAZAAALPONDRLDPEAYGAPCARSGQPMQVSLFNGLSFHCAGV 74  
 DB 2 RAL-----LPLALVGAAYV--FPVDD--DKIVGYTCQNSVPYQVSLNAGYHF--CGGS 51  
 QY 75 LVDSQSWILTAHCGNKPMARVGDHLLILQG--EQLRRTTSVVPKXHQSGPILPRRT 133  
 DB 52 LINDQWVSAACIKYKSGRIQVRLGEHNINVLGEDEQFINAKIKHFV-----DRKT 103

OY	134	DEHIMLMKARVAVGEPFRBALQIYRCAGPBGDCCYAAGCTTAARVKYNKLTCSSI	193
Dd	104	LANDMIMIKISSPEVKLNARVATALPSSCAPAGTCLLSIGMGNTLSSGVNEPDLLQCIDA	163
OY	194	TILSPKECEVFYEGVVTTNNNICAG-LDRQDDPCGSDSGLVCBETLOGILLSMEVYCGS	253
Dd	164	PLLPQADCEASTGYEKTIDNNVCCGFLLEGKSCQSDSGSPVVGNGELGIVSMG-YGCAL	222
OY	253	AQHPAVYTQICRYMSWINKVRNS	276
Dd	223	PDNFGVTTKVCNVDMIQDTIAAN	246

RESULT 8  
S05494  
trypsin (BC 3.4.21.4) IV precursor - rat  
N:Alternate names: 23K protein; trypsinogen IV precursor  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 22-Jun-1999  
C:Accession: S05494  
R:Linette, H.; Rausch, V.; Vasiloudes, P.; Scheele, G.A.; Kern, H.F.  
Nucleic Acids Res. 17, 6736, 1989  
A:Title: A fourth trypsinogen (P23) in the rat pancreas induced by CCK.  
A:Reference number: S05494; MUID:89386010; PMID:2780302  
A:Accession: S05494  
A:Molecule type: mRNA  
A:Residues: 1-247 <IUE>  
A:Cross-references: EMBL:X15679; NID:956813; PIRN:CA03778.1; PID:956814  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: calcium binding; hydrolase; protein digestion; serine proteinase; zymogen  
F:1-15/Domain: signal sequence #status predicted <SIG>  
F:16-23/Domain: activation peptide #status predicted <APT>  
F:24-240/Product: trypsin IV #status predicted <MAT>  
F:24-240/Domain: trypsin homology <TRY>  
F:30-161,49-65,133-224,140-207,172-186/Disulfide bonds: #status predicted  
F:64,108,201/Active site: His, Asp, Ser #status predicted  
F:76,78,81,86/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query March	32.5%;	Score 486.5;	DB 2;	Length 247;
Best Local Similarity	39.5%;	Pred. No. 4.7e-35;		
Matches 102;	Conservative 44;	Mismatches 97;	Indels 15;	Gaps 6

QY	2	LEPLMAQVLAAEKALLPONDTRLDEBAVGAFCARBSQPMQVSLFNGLSFHOAGVLYVQSW	80
Db	3	ISIFFFFLAANA-VA--LPVND--DKIVGGVTPKPLVPEYVSLSHGIGSHQGGSLISDQM	58
QY	81	VLTAAHCGRKPMARVGDHLLTLQ-GEQRTETTSVHPKXQSGSLPRFRDEHDM	139
Db	59	VLSAAHCYKRLQVRLGEGNHIVLPGSGGFIDAEKIIHPERN-----KOTLDNDIM	110
QY	140	LTKLAPVVPSPRYALOLPYRCAPRSGOCVAGSGCTAARRYKYNKGLTSSITIIISPK	199
Db	111	LTKLSPVAVNSQVSVSLPRSCASTDACLVSQGNVTASVGGKRPALLQCLAPVYLSAS	170
QY	200	ECGEVFPFQVYVNNMTCAG-LDRGQDPCCSDSGSLVCDETLQGLISWGVPCGSAQHPAY	256
Db	171	SCCKASIPGGITSNMFGLFIBGCKDSCDSDSGPVYCNGETIGIIVSGSV-CANRGKPGV	229
QY	259	YVQICKYMSWINKVIRSN	276
Db	230	YTKVCNVLISWIGETMANN	247

RESULT 9  
TRPGTR  
N:crystin (BC 3\_4.21.4) precursor - pig (tentative sequence)  
C:Species: Sus scrofa domestica (domestic pig)  
C>Date: 24-Apr-1984 #sequence\_rev:150 #text\_change 31-Mar-2000  
R:Accession: A90641; A90360; A00947  
R:Charles, M.; Roversy, M.; Guldorf, A.; Desmuelle, P.  
Biochim. Biophys. Acta 69, 115-129, 1963

A/Title: Su le trypsinogene et la trypsine de porc.  
 2: Reference number: 200647

A;Reference number: A90641

A;Molecule type: protein

R; Hermodson, M.A.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.

**A; Title:** Determination of the amino acid sequence of p

A;Reference number: A90368; MUID:73258692; PMID:47389333  
A;Accession: A90368

A;Molecule type: protein

A:Note: at position 20, Ile and Val occur alternatively

**Keywords:** hydrolase; pancreas; polymorphism; protein digestion; product; trypsinogen #status experimental; F-1-231/

```
F;1-8/Domain: activation peptide #status experimental <APT:
E:9-231/Product: trypsin #status experimental -MM:

```

E:15-145\_33-49\_117-218\_124-191156-170\_181-205/D1seq1.fide bo

E	48,92,185	Active site: His, Asp, Ser	#status predicted
E:60,62,65-70	/Binding site: calcium (Ca <sup>2+</sup> )	Asn Val Gly	#status predicted

Query Match	32.38;	Score 483;	DB 1;	Length 231;
-------------	--------	------------	-------	-------------

Matches 93; Conservative 41; Mismatches 84; Indels 12; Gaps 5

49 GAPCARGSQFWQVSLFNLSPFHCAGVLVDQSWLTAHCGNKPIMARVGDHLLLLQG-E 107

12 GYTCAANSIPYQVSLNSGSHF-CGSLINSQWVSAHCHYKSRIQVRLGEHNIDVLEONE 70

QY 108 QLRRTTRSVVHPKYNHGGSPILPRTDEHDLMLKLARPVVEGPRVRALQLPYRCAQPGD 167

Db 71 QFINAKIYTHNEN-----GNTLDNDIMLIKSSPATLNSRVATVSLPRSCAAGT 122

168 QCGVAGWGTTAARRVKYNNKGLTSSITILSPKECEVFYPGVVTNNMICAG-LDRGQDPQ 226

123 ECLISGWGNTKSSGSSYP<sup>SL</sup>LQCLKAPVLSDDSSCKSSYPGQITGNMIVGFL<sup>EG</sup>GKDS<sup>CQ</sup> 182

227 SDSGPLVCDETLQGILSNQVYPCGSAQHPAVYTOICKYMSWINKVIRSN 276

Db 183 GDSGPRVNCNGQLQGIIVSWG-YGCAQKNKPGVYTKVCNYYVNWIIQQTIAAN 231

RESULT 10  
TRRT1

trypsin (EC 3.4.21.4) I precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C/Accession: B22657; A00948

J. Biol. Chem. 259, 14255-14264, 1984

A:Reference number: A22657; MUID:85054880; PMID:6094547  
A:Accession: B32657

A: Molecule type: DNA  
A: Residues: 1-246 <CPA>

A:Note: the authors translated the codon ATC for position 5 in low and high calcium

K; MacDonald, R.J.; Staley, S.J.; Switt, G.H.  
J. Biol. Chem. 257: 9724-9732, 1982

A/Reference number: A00948; MUID:82265624; PMID:6896710

A/Molecule type: mRNA

A;Cross-references: GB:J00778: NID:q206507: PIDN:AAA98518.1: PTD:q206508  
r/ncvalues: 1-248 <MACS

A;introns: 14/1; 67/2; 152/1; 197/3

C;Keywords: hydrolase; pancreas: serine proteinase: zymogen

F;1-15/Domain: signal sequence #status predicted <SIG>

F:16-23/Domain: activation peptide #status predicted <APT>  
 F:24-26/Product: trypsin I #status predicted <ENZ>  
 F:24-23/Domain: trypsin homology <TRY>  
 F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted  
 F:63,107,200/Active site: His, Asp, Ser #status predicted  
 F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 32.2%; Score 482; DB 1; Length 246;  
 Best Local Similarity 39.5%; Pred. No. 1,2e-34;  
 Matches 101; Conservative 44; Mismatches 95; Indels 16; Gaps 6;

QY 23 LLMQAQMAEALPNDRLDPEAYGACGASQPMQVSLFNGLSFHCAGVLVDQSVYL 82  
 D 5 LILALGAVALPELDDDKIVG---GYCPEHSIPYQVSLNSGYNHF-CGSLINDQVAV 59  
 QY 83 TAAHCGKRLMARVGDHLLILQ-EGLRRTSRVHPKXGSGPILPRTRDEHML 141  
 D 60 SAHCKYKSRIOVRLGSHNINVLGEQPIINAKIKHPYSSM-----TLNNDIMLI 111  
 QY 142 KLARPVVPGRRVALQTPRCAPQPGQCVAGMGTTAARVRYKNGLTCSITLSPKEC 201  
 D 112 KLSPPVKLMARVAPVLPSCAPAGNQCULSGMNTLSGVNPPDLQCVAPVLSQADC 171  
 QY 202 EVFPEVATNNMICAG-LDRGDDPCQSDSGPLVCDTLOGLISGVVPCGSAQHPAYT 260  
 D 172 EAAVPELITSMICVGFLEGGKSDSCSDSGPVCNGQLOGLIVSWG-YGCALEPDPGYT 230  
 QY 261 QICKYMSINKVIRSN 276  
 D 231 KVCNFGMIQDTIAAN 246

## RESULT 11

A27547  
 trypsin (EC 3.4.21.4) precursor, cationic - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 22-Jun-1999  
 C/Accession: A27547  
 R/Release: T.S.; Alhadeff, M.; Craik, C.S.; Laryman, C.  
 Biochemistry 26, 3081-3086, 1987  
 A/Title: Isolation and characterization of a cDNA encoding rat cationic trypsinogen.  
 A/Reference number: A27547; MUID:8721609; PMID:3607011  
 A/Accession: A27547  
 A/Molecule type: mRNA  
 A/Residues: 1-247 <ENB>  
 A/Cross-references: GB:M2624; NID:9206498; PIDN:AAA1985.1; PID:9206499  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: calcium binding; hydrolase; protein digestion; serine proteinase  
 F:25-240/Domain: trypsin homology <TRY>  
 F:31-161,49-65,133-234,140-207,172-186/Disulfide bonds: #status predicted  
 F:64,108,201/Active site: His, Asp, Ser #status predicted  
 F:76,78,81,86/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 31.9%; Score 477.5; DB 2; Length 247;  
 Best Local Similarity 37.5%; Pred. No. 2.9e-34;  
 Matches 96; Conservative 47; Mismatches 98; Indels 15; Gaps 6;

QY 23 LLMQAQMAEALPNDRLDPEAYGACGASQPMQVSLFNGLSFHCAGVLVDQSVYL 82  
 D 5 LILALGAVALPELDDDKIVG---GYCPEHSIPYQVSLNSGYNHF-CGSLINDQVAV 60  
 QY 83 TAAHCGKRLMARVGDHLLILQ-EGLRRTSRVHPKXGSGPILPRTRDEHML 141  
 D 61 SAHCKYKSRIOVRLGSHNINVLGEQPIINAKIKHPYSSM-----TLNNDIMLI 112  
 QY 142 KLARPVVPGRRVALQTPRCAPQPGQCVAGMGTTAARVRYKNGLTCSITLSPKEC 201  
 D 113 KLSPPVKLMARVAPVLPSCAPAGNQCULSGMNTLSGVNPPDLQCVAPVLSQADC 172  
 QY 202 EVFPEVATNNMICAG-LDRGDDPCQSDSGPLVCDTLOGLISGVVPCGSAQHPAYT 260  
 D 173 KSVYPRKITSNMFCLGLEGKSDSCSDSGPVCNGQLOGLIVSWG-YGCALEPDPGYT 231

QY 261 QICKYMSINKVIRSN 276  
 D 232 KVCNFGMIQDTIAAN 247

## RESULT 12

S49489  
 trypsin (EC 3.4.21.4) precursor - Paratocthenia magellanica  
 C/Species: Paratocthenia magellanica  
 C/Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 22-Jun-1999  
 A/Accession: S49489  
 R/Gene: S. Renteria-Deluna, F. Edwards, D. van Beunnen, J. Dodson, G. Gerday, C  
 submitted to the EMBL Data Library, October 1994  
 A/Description: Trypsin and trypsinogen from an antarctic fish: molecular basis of cold  
 A/Reference number: S49489  
 A/Accession: S49489  
 A/Molecule type: mRNA  
 A/Residues: 1-242 <GBN>  
 A/Cross-references: EMBL:X82223; NID:9559507; PIDN:CAA57701.1; PID:9559508  
 C/Superfamily: trypsin; trypsin homology  
 C/Keywords: hydrolase; serine proteinase  
 F:1-13/Domain: signal sequence #status predicted <SIG>  
 F:14-242/Product: trypsin #status predicted <MAT>  
 F:21-235/Domain: trypsin homology <TRY>

Query Match 31.7%; Score 473.5; DB 2; Length 242;  
 Best Local Similarity 41.1%; Pred. No. 6.3e-34;  
 Matches 95; Conservative 43; Mismatches 80; Indels 13; Gaps 6;

QY 41 TRIDPEAYGAPCARSGQPMQVSLFNGLSFHCAGVLVDQSVYLAHCGKRLMARVGDH 100  
 D 16 TEBDKIVGKESCPYQPHQVSLNSGYHF-CGSLINENMVVSAHCKYKSRVREVRGEHH 74  
 QY 101 LLLLOG-EGLRRTSRVHPKXGSGPILPRTRDEHMLKLARPVVPGRRVALQ 159  
 D 75 IRTYBKGEGRTSSRRVIRHPNYS-----YINDNDIMLIKLSRPAALNQVQVALP 126  
 QY 160 YRCAPQDQCVAGMGTTAARVRYKNGLTCSITLSPKECVFFGVATNNMICAG-L 218  
 D 127 SSCAPAGTCTVSGMSTQSSADGNK-LQCLNIPILSRDCDNSYPMITDAMPAGYL 185  
 QY 219 DRGDDPCQSDSGPLVCDTLOGLISGVVPCGSAQHPAYTQICKYMSNI 269  
 D 186 QGGKSDSCSDSGPVCNGQLOGLIVSWG-YGCALEPDPGYAKVCLFNDWL 235

## RESULT 13

B25852  
 trypsin (EC 3.4.21.4) II precursor (validated) - human  
 N/Alternate names: trypsin 2; trypsin, anionic; trypsinogen II  
 C/Species: Homo sapiens (man)  
 C/Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 08-Dec-2000  
 C/Accession: B25852; A61066; B43988  
 R/Emt, M.; Nakamura, Y.; Ogawa, M.; Yamamoto, T.; Nishide, T.; Mori, T.; Matsubara, K.  
 Gene 41, 305-310, 1986  
 A/Title: Cloning, characterization and nucleotide sequences of two cDNAs encoding huma  
 A/Reference number: A91544; MUID:86221712; PMID:3011602  
 A/Accession: B25852  
 A/Molecule type: mRNA  
 A/Residues: 1-247 <EMI>  
 A/Cross-references: GB:M27602; NID:9521217; PIDN:AAA61232.1; PID:9521218  
 R/Kimland, M.; Russick, C.; Marks, W.H.; Borgstrom, A.  
 Clin. Chim. Acta 184, 31-46, 1989  
 A/Title: Immunoreactive anionic and cationic trypsin in human serum.  
 A/Reference number: A61066; MUID:90091010; PMID:2598466  
 A/Accession: A61066  
 A/Molecule type: protein  
 A/Residues: 16-39, 'X', 41-42, 'XXX', 47-49 <KIM>  
 R/Koivunen, E.; Huhtala, M.L.; Stenman, U.H.  
 J. Biol. Chem. 264, 14095-14099, 1989  
 A/Title: Human ovarian tumor-associated trypsin. Its purification and characterization  
 A/Reference number: A43988; MUID:89340515; PMID:2503510  
 A/Accession: B43988

A.Molecule type: protein  
 A.Residues: 16-49 <KOI>  
 A.Experimental source: mucinous ovarian tumor cyst fluid  
 C.Genetics:  
 A.Gene: GDB:P8582; TRX2  
 A.Cross-references: GDB:335289; OMIM:601564  
 A.Map position: 7q35-7q35  
 A.Note: The human genome contains at least ten trypsin genes or pseudogenes, at least two  
 C.Superfamily: trypsin; trypsin homology  
 C.Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
 F.1-15/Domain: signal sequence #status predicted <SIG>  
 F.16-246/Product: trypsinogen II #status experimental <ZYM>  
 F.16-23/Domain: activation peptide #status predicted <ENM>  
 F.24-246/Product: trypsin II #status predicted <ENM>  
 F.24-239/Domain: trypsin homology <TRY>  
 F.30-160,48-64,171-185,196-220/Disulfide bonds: #status predicted  
 F.63,107,200/Active site: His, Asp, Ser #status predicted  
 F.75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 31.4%; Score 470; DB 1; Length 247;  
 Best Local Similarity 38.7%; Pred. No. 1,3e-33;  
 Matches 99; Conservative 41; Mismatches 102; Indels 14; Gaps 6;

QY 23 LLMQLMAAEAPALLPQNDTRIDPEAYGAPCARSGQPMQVSLFNGLSFHCAGLVLDQSWL 82  
 Db LLLILTFVAAVAAPFDD--DKIVGVCENSVPEVQSLNSGYHF--CGSLSISQWV 59  
 QY 83 TAAOCNKLMAEYGDHLLILQ--EQLRRTTSVVPKXGSGPILPRDEHML 141  
 Db SAGHYKSRIOVRLGEHNIENLESGEPTNAKIRPKNS-----ETLNDIL 111  
 QY 142 KLARPVVGFPRVRLQLPYRCAPGQDQVAGWGTAAARVKNKGLTSSITILSKPC 201  
 Db 112 KLSPAVINSHVSAISLPTAPPAAGTSLISGMNLTSSADVPDELOCDAPVLSQAC 171  
 QY 202 EVFPGVVTNMICAG--LDRGDPCCSDSGPLVCDETLGSLISWGVPCGSAQHVAVT 260  
 Db 172 EASIFGKTIINMFCVGLGKDCSCQDSGSPVSNBLQITVSN--YGCAGKRRGVIT 230  
 QY 261 QICKYMSINKVIRSN 276  
 Db 231 KYNYVDIKDTIAAN 246

## RESULT 14

SS066  
 trypsin (EC 3.4.21.4) II precursor, pancreatic (clone 2-P29) - chicken  
 N.Alternate names: trypsinogen II  
 C.Species: Gallus gallus (chicken)  
 C.Date: 23-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 22-Jun-1995  
 C.Accession: S55066; S72347  
 R.Wang, K.; Gan, L.; Lee, I.; Hood, L.  
 Biochem. J. 307, 471-479, 1995  
 A>Title: Isolation and characterization of the chicken trypsinogen gene family.  
 A.Reference number: S55065; MUID:95251611; PMID:7733885  
 A.Accession: S55066  
 A.Molecule type: mRNA  
 A.Residues: 1-248 <MAN1>  
 A.Cross-references: EMBL:U15157; NID:G603906; PIDN:AAA79914.1; PID:G603907  
 A.Experimental source: clone 2-P29  
 A.Accession: S72347  
 A.Molecule type: DNA  
 A.Residues: 1-248 <MAN2>  
 A.Cross-references: EMBL:U15157; NID:G603906; PIDN:AAA79914.1; PID:G603907  
 A.Experimental source: clone 2-P29  
 C.Superfamily: trypsin; trypsin homology  
 C.Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen  
 F.1-16/Domain: signal sequence #status predicted <SIG>  
 F.17-25/Domain: activation peptide #status predicted <APT>  
 F.26-248/Product: trypsin II #status predicted <MAN>  
 F.26-241/Domain: trypsin homology <TRY>  
 F.65,109,202/Active site: His, Asp, Ser #status predicted

Query Match 31.4%; Score 469.5; DB 2; Length 248;  
 Best Local Similarity 38.5%; Pred. No. 1.4e-33;  
 Matches 99; Conservative 52; Mismatches 91; Indels 15; Gaps 6;

QY 19 KLPLMAQMAEAPALLPQNDTRIDPEAYGAPCARSGQPMQVSLFNGLSFHCAGLVLDQ 78  
 Db 2 KFLPLISCLGANAFAFGADDDKI---VGGYCPHNSVYQVLSNGHF--CGSLSINS 57  
 QY 79 SWVLAACGNKPLMAEYGDHLLILQEQLRRTTSV--HPKXGSGPILPRDEHD 137  
 Db 58 QWVLSAHCYSRIQVRLGEYNIIDVQDSSEVRSVILIRPKYS-----ITLND 109  
 QY 138 LMLKLAPVPGPRVRLQLPYRCAPGQDQVAGWGTAAARVKNKGLTSSITILS 197  
 Db 110 LMLKLASVYSDIDPILPSSCAAGTECLISGMNLTSSADVPDELOCDAPVLSQAC 169  
 QY 198 PKECEVPGVVTNMICAG--LDRGDPCCSDSGPLVCDETLGSLISWGVPCGSAQHP 256  
 Db 170 DQCEAVPGDITSNMFCVGLGKDCSCQDSGSPVSNBLQITVSN--YGCAGKRRGVIT 228  
 QY 257 AVYQICKYMSINKVI 273  
 Db 229 GVTYKVCNVDIODETI 245

## RESULT 15

SS1779  
 trypsin (EC 3.4.21.4) III precursor - Atlantic salmon (fragment)  
 C.Species: Salmo salar (Atlantic salmon)  
 C.Date: 22-Nov-1995 #sequence\_revision 03-Aug-1995 #text\_change 22-Jun-1999  
 C.Accession: S66657; S31779  
 R.Male, R.; Lorens, J.B.; Smals, A.O.; Torrisen, K.R.  
 Eur. J. Biochem. 232, 677-685, 1995  
 A>Title: Molecular cloning and characterization of anionic and cationic variants of trypsin.  
 A.Reference number: S66657; MUID:9603908; PMID:7556223  
 A.Accession: S66657  
 A.Molecule type: mRNA  
 A.Residues: 1-238 <MAL>  
 A.Cross-references: EMBL:X70074; NID:G64387; PIDN:CAA49679.1; PID:G64388  
 C.Superfamily: trypsin; trypsin homology  
 C.Keywords: hydrolase; signal sequence (fragment) #status predicted <SIG>  
 F.1-7/Domain: signal sequence (fragment)  
 F.8-15/Domain: activation peptide #status predicted <APT>  
 F.16-238/Product: trypsin III #status predicted <MAN>  
 F.16-231/Domain: trypsin homology <TRY>  
 F.22-152,40-56,124-225,131-198,163-177,188-212/Disulfide bonds: #status predicted  
 F.55,99,192/Active site: His, Asp, Ser #status predicted

Query Match 31.3%; Score 468.5; DB 2; Length 238;  
 Best Local Similarity 37.5%; Pred. No. 1.7e-33;  
 Matches 93; Conservative 46; Mismatches 94; Indels 15; Gaps 6;

QY 31 AEAALLPQNDTRIDPEAYGAPCARSGQPMQVSLFNGLSFHCAGLVLDQSWLTAACGNK 90  
 Db 4 AEAAPIDDEDKI---VGGYECRGRSASYSQSLQSGHF--CGSLSISSTWVSAHCYKS 59  
 QY 91 PLMARVGDHLLILQ--EQLRRTTSVVPKXGSGPILPRDEHMLKLARPVV 149  
 Db 60 RIQVRLGSHNIAVNBGRQFLDSVKNIMHRYNS-----RLUNDILKLSKRSAL 111  
 QY 150 GPRVALDLPYRCAPGQDQVAGWGTAAARVKNKGLTSSITILSPKECEVPGV 209  
 Db 112 NSVYSTVALPSSCASGGRCLVSGMNLSSGSSNYPTLCLDPLILSSSSCASAPGQI 171  
 QY 210 TNNMTCAG--LDRGDPCCSDSGPLVCDETLGSLISWGVPCGSAQHPAVYTOICKYMSW 268  
 Db 172 TSNMFCAGFMGKADSCQDSGSPVSNBLQITVSN--YGCAGKRRGVITTKCYNRSW 230  
 QY 269 INKTVIRSN 276  
 Db 231 ISSTWSSN 238



Wed Nov 26 18:23:54 2003

us-10-021-368-1.rpt

Page 7

Search completed: November 25, 2003, 09:09:13  
Job time : 28 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 09:08:46 | Search time 45 Seconds

(without alignments)  
1131.252 Million cell updates/sec

Title: US-10-021-368-1

Perfect score: 1496  
Sequence: 1 MRAPHILHLSAASGARALATL.....AVYTOICKYMSINKVIRSN 276

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 673684 seqs, 184443283 residues

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications AA:\*

1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
9: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/prodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
11: /cgn2\_6/prodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/prodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
13: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/prodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/prodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
17: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/prodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1496	100.0	276	9	US-09-888-615-87
2	1496	100.0	276	12	US-10-301-822-95
3	1496	100.0	276	12	US-10-366-288-52
4	1496	100.0	276	14	US-10-021-368-1
5	1496	100.0	276	15	US-10-097-340-167
6	1438.5	96.2	291	14	US-10-021-368-11
7	564	37.7	248	11	US-09-946-374-194
8	564	37.7	248	12	US-10-013-387A-194
9	564	37.7	248	12	US-10-006-130A-194
10	564	37.7	248	12	US-10-006-172A-194
11	564	37.7	248	12	US-10-015-392A-194
12	564	37.7	248	12	US-10-017-253A-194
13	564	37.7	248	12	US-10-017-306A-194
14	564	37.7	248	12	US-10-012-064A-194
15	564	37.7	248	12	US-10-017-867A-194

16	564	37.7	248	12	US-10-012-101B-194	Sequence 194, App
17	564	37.7	248	12	US-10-012-137A-194	Sequence 194, App
18	564	37.7	248	12	US-10-012-752A-194	Sequence 194, App
19	564	37.7	248	12	US-10-012-754A-194	Sequence 194, App
20	564	37.7	248	12	US-10-013-809A-194	Sequence 194, App
21	564	37.7	248	12	US-10-013-911A-194	Sequence 194, App
22	564	37.7	248	12	US-10-013-912A-194	Sequence 194, App
23	564	37.7	248	12	US-10-015-610A-194	Sequence 194, App
24	564	37.7	248	12	US-10-015-653A-194	Sequence 194, App
25	564	37.7	248	12	US-10-015-671A-194	Sequence 194, App
26	564	37.7	248	12	US-10-012-237A-194	Sequence 194, App
27	564	37.7	248	12	US-10-013-306A-194	Sequence 194, App
28	564	37.7	248	12	US-10-015-386A-194	Sequence 194, App
29	564	37.7	248	12	US-10-015-480A-194	Sequence 194, App
30	564	37.7	248	12	US-10-015-715A-194	Sequence 194, App
31	564	37.7	248	12	US-10-012-753A-194	Sequence 194, App
32	564	37.7	248	12	US-10-015-385A-194	Sequence 194, App
33	564	37.7	248	12	US-10-007-236A-194	Sequence 194, App
34	564	37.7	248	12	US-10-015-389A-194	Sequence 194, App
35	564	37.7	248	12	US-10-013-915A-194	Sequence 194, App
36	564	37.7	248	12	US-10-015-394A-194	Sequence 194, App
37	564	37.7	248	12	US-10-015-519A-194	Sequence 194, App
38	564	37.7	248	12	US-10-015-390A-194	Sequence 194, App
39	564	37.7	248	15	US-10-006-856A-194	Sequence 194, App
40	564	37.7	248	15	US-10-006-818A-194	Sequence 194, App
41	564	37.7	248	15	US-10-015-393A-194	Sequence 194, App
42	564	37.7	248	15	US-10-015-869A-194	Sequence 194, App
43	564	37.7	248	15	US-10-012-121A-194	Sequence 194, App
44	564	37.7	248	15	US-10-006-116A-194	Sequence 194, App
45	564	37.7	248	15	US-10-006-117A-194	Sequence 194, App

#### ALIGNMENTS

RESULT 1  
US-09-888-615-87  
Sequence 87, Application US/0988615  
Patent No. US2002064856A1  
GENERAL INFORMATION:  
APPLICANT: FLOWMAN, GREGORY  
APPLICANT: WHITE, DAVID  
APPLICANT: CARRISPEL, SEAN  
APPLICANT: CHARYCZAK, GLEN  
APPLICANT: MANNING, GERRARD  
APPLICANT: SUDASANAM, SUCHA  
TITLE OF INVENTION: NOVEL PROTEASES  
FILE REFERENCE: 038602/1214  
CURRENT APPLICATION NUMBER: US/09/888,615  
CURRENT FILING DATE: 2001-06-26  
PRIOR APPLICATION NUMBER: 60/214,047  
PRIOR FILING DATE: 2000-06-26  
NUMBER OF SEQ ID NOS: 150  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 87  
LENGTH: 276  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-888-615-87

Query Match 100.0%; Score 1496; DB 9; Length 276;  
Best Local Similarity 100.0%; Pred. No. 5.7e-141;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRAPHILHLSAASGARALATLPLMAQMAAALIPONTTLPDPAAGAPARSSQWQ	60
Db	1	MRAPHILHLSAASGARALATLPLMAQMAAALIPONTTLPDPAAGAPARSSQWQ	60
Qy	61	VSLENGLSFPCAGAVLDQSWVLTAAHCGNKPPLMARVGDHLLLOEOQLRTRTVVHPK	120
Db	61	VSLENGLSFPCAGAVLDQSWVLTAAHCGNKPPLMARVGDHLLLOEOQLRTRTVVHPK	120
Qy	121	YHGGSGILPRTTDEHMLKLARPVVPGRVRLQLPYRCAPGDCQVAGMKTAAAR	180

Db 121 YHOGSGPILPRTDEHDLMLKLARPVVPGPRVALQLPYRCAGQGDQCQVAGMGTAR 180  
QY 181 RYKYNKGLTSSITILSPKECEVEFYGVVNTNMICAGLDRGDDPCQSDSGGLVCDETLQ 240  
Db 181 RYKYNKGLTSSITILSPKECEVEFYGVVNTNMICAGLDRGDDPCQSDSGGLVCDETLQ 240  
QY 241 GILSMGVYPCGSAQHPAVYTOICKYMSWINKYIRSN 276  
Db 241 GILSMGVYPCGSAQHPAVYTOICKYMSWINKYIRSN 276

## RESULT 2

US-10-301-822-95  
; Sequence 95, Application US/10301822  
; Publication No. US20030148410A1  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc.  
; APPLICANT: Berger, Allison  
; APPLICANT: Guillemette, Tracy L.  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Thibodeau, Stephen N.  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; FILE REFERENCE: MPM01-029P2RNM  
; CURRENT APPLICATION NUMBER: US/10/301,822  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 60/339,971  
; PRIOR FILING DATE: 2001-12-10  
; PRIOR APPLICATION NUMBER: US 60/361,978  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 60/381,988  
; PRIOR FILING DATE: 2002-05-20  
; NUMBER OF SEQ ID NOS: 228  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 95  
; LENGTH: 276  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-301-822-95

Query Match 100.0%; Score 1496; DB 12; Length 276;  
Best Local Similarity 100.0%; Pred. No. 5.7e-141;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAPHLHLSAASGARALAKLPLMLMAQUMAERALLPQNDTRLDEPAVAGAPCARSGSPWQ 60  
Db 1 MRAPHLHLSAASGARALAKLPLMLMAQUMAERALLPQNDTRLDEPAVAGAPCARSGSPWQ 60  
QY 61 VSLFNGLSFHCAGVLVDQSWVLTAAHCGNKLPMARVGDHLLLOEQLRRTTTSVHPK 120  
Db 61 VSLFNGLSFHCAGVLVDQSWVLTAAHCGNKLPMARVGDHLLLOEQLRRTTTSVHPK 120  
QY 121 YHOGSGPILPRTDEHDLMLKLARPVVPGPRVALQLPYRCAGQGDQCQVAGMGTAR 180  
Db 121 YHOGSGPILPRTDEHDLMLKLARPVVPGPRVALQLPYRCAGQGDQCQVAGMGTAR 180  
QY 181 RYKYNKGLTSSITILSPKECEVEFYGVVNTNMICAGLDRGDDPCQSDSGGLVCDETLQ 240  
Db 181 RYKYNKGLTSSITILSPKECEVEFYGVVNTNMICAGLDRGDDPCQSDSGGLVCDETLQ 240  
QY 241 GILSMGVYPCGSAQHPAVYTOICKYMSWINKYIRSN 276  
Db 241 GILSMGVYPCGSAQHPAVYTOICKYMSWINKYIRSN 276

RESULT 3  
US-10-366-288-52  
; Sequence 52, Application US/10366288

Publication No. US20030216288A1

; GENERAL INFORMATION:  
; APPLICANT: Powell, Douglas  
; APPLICANT: Welch, Nadine S.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING  
; TITLE OF INVENTION: AIDS AND HIV-RELATED DISORDERS USING 1414, 1481, 1553,  
; TITLE OF INVENTION: 34021, 1720, 1683, 1552, 1682, 12825, 9952, 5816,  
; TITLE OF INVENTION: 10002, 1611, 1371, 14324, 126, 270, 312, 167, 326, 18926,  
; TITLE OF INVENTION: 6747, 1793, 1784 OR 2045 MOLECULES  
; FILE REFERENCE: MP102-025P1RNMNMIM  
; CURRENT APPLICATION NUMBER: US/10/366,288  
; CURRENT FILING DATE: 2003-02-13  
; PRIOR APPLICATION NUMBER: 60/357,391  
; PRIOR FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 60/380,249  
; PRIOR FILING DATE: 2002-05-13  
; PRIOR APPLICATION NUMBER: 60/391,306  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: 60/406,297  
; PRIOR FILING DATE: 2002-08-27  
; PRIOR APPLICATION NUMBER: 60/412,007  
; PRIOR FILING DATE: 2002-09-19  
; PRIOR APPLICATION NUMBER: 60/417,508  
; PRIOR FILING DATE: 2002-10-10  
; PRIOR APPLICATION NUMBER: 60/432,318  
; PRIOR FILING DATE: 2002-12-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 52  
; LENGTH: 276  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-366-288-52

Query Match 100.0%; Score 1496; DB 12; Length 276;  
Best Local Similarity 100.0%; Pred. No. 5.7e-141;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAPHLHLSAASGARALAKLPLMLMAQUMAERALLPQNDTRLDEPAVAGAPCARSGSPWQ 60  
Db 1 MRAPHLHLSAASGARALAKLPLMLMAQUMAERALLPQNDTRLDEPAVAGAPCARSGSPWQ 60  
QY 61 VSLFNGLSFHCAGVLVDQSWVLTAAHCGNKLPMARVGDHLLLOEQLRRTTTSVHPK 120  
Db 61 VSLFNGLSFHCAGVLVDQSWVLTAAHCGNKLPMARVGDHLLLOEQLRRTTTSVHPK 120  
QY 121 YHOGSGPILPRTDEHDLMLKLARPVVPGPRVALQLPYRCAGQGDQCQVAGMGTAR 180  
Db 121 YHOGSGPILPRTDEHDLMLKLARPVVPGPRVALQLPYRCAGQGDQCQVAGMGTAR 180  
QY 181 RYKYNKGLTSSITILSPKECEVEFYGVVNTNMICAGLDRGDDPCQSDSGGLVCDETLQ 240  
Db 181 RYKYNKGLTSSITILSPKECEVEFYGVVNTNMICAGLDRGDDPCQSDSGGLVCDETLQ 240  
QY 241 GILSMGVYPCGSAQHPAVYTOICKYMSWINKYIRSN 276  
Db 241 GILSMGVYPCGSAQHPAVYTOICKYMSWINKYIRSN 276

RESULT 4  
US-10-021-368-1  
; Sequence 1, Application US/10021368  
; Publication No. US20020106367A1  
; GENERAL INFORMATION:  
; APPLICANT: Band, Yvonne  
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
; MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA

COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/021,368  
FILING DATE: 12-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 09/201,038  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00398/100002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 276 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-021-368-1

Query Match 100.0%; Score 1496; DB 14; Length 276;  
Best Local Similarity 100.0%; Pred. No. 5,7e-141;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAPHLHLSAAGBARALAKLPLMLQMAAEALPLPNDTRLDPEAYGAPCARSGOPMQ 60  
DB 1 MRAPHLHLSAAGBARALAKLPLMLQMAAEALPLPNDTRLDPEAYGAPCARSGOPMQ 60  
QY 61 VSLFNGSLFPCAGVLDQSVLTAACGKPKLWAVGDHLLLLQGEQLRTTRSVVHPK 120  
DB 61 VSLFNGSLFPCAGVLDQSVLTAACGKPKLWAVGDHLLLLQGEQLRTTRSVVHPK 120  
QY 121 YHSGSGILPRTRDEHMLKLARPVPGPRVRAQLPYRCAPDPDCCQVAGMGTTAAR 180  
DB 121 YHSGSGILPRTRDEHMLKLARPVPGPRVRAQLPYRCAPDPDCCQVAGMGTTAAR 180  
QY 181 RYKYNKGLTCSSTITLSPKCEVFYFGVVTNNMICAGLDKGDPCQSDSGPLVCDETIQ 240  
DB 181 RYKYNKGLTCSSTITLSPKCEVFYFGVVTNNMICAGLDKGDPCQSDSGPLVCDETIQ 240  
QY 241 GILSWGVPDGSAGHPAVYTQICKYMSWINKVIRSN 276  
DB 241 GILSWGVPDGSAGHPAVYTQICKYMSWINKVIRSN 276

RESULT 5  
US-10-097-340-167  
Sequence 167, Application US/10097340  
Publication No. US2003087250A1  
GENERAL INFORMATION:

APPLICANT: John MONAHAN  
APPLICANT: Manjula GANNAVARAPU  
APPLICANT: Sebastian HOERSCHE  
APPLICANT: Shubhangi KAMATKAR  
APPLICANT: Steve G. KOVATS  
APPLICANT: Rachel E. MEYERS  
APPLICANT: Michael MORRISSEY  
APPLICANT: Peter OLANDT  
APPLICANT: Ami SEN  
APPLICANT: Peter VEIBY  
APPLICANT: Gordon B. MILLS

APPLICANT: Robert C. BAST, Jr.  
APPLICANT: Karen LU  
APPLICANT: Rosemarie SCHMANDT  
APPLICANT: Xumei ZHAO  
APPLICANT: Karen GLATT  
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,  
Assessment, Prevention, and Therapy Of Ovarian Cancer  
FILE REFERENCE: MRI-030  
CURRENT APPLICATION NUMBER: US/10/097,340  
PRIORITY FILING DATE: 2002-03-14  
PRIORITY APPLICATION NUMBER: 60/276,025  
PRIORITY FILING DATE: 2001-03-14  
PRIORITY APPLICATION NUMBER: 60/325,149  
PRIORITY FILING DATE: 2001-09-26  
PRIORITY APPLICATION NUMBER: 60/276,026  
PRIORITY FILING DATE: 2001-03-14  
PRIORITY APPLICATION NUMBER: 60/324,967  
PRIORITY FILING DATE: 2001/09/26  
PRIORITY APPLICATION NUMBER: 60/311,732  
PRIORITY FILING DATE: 2001-08-10  
PRIORITY APPLICATION NUMBER: 60/325,102  
PRIORITY FILING DATE: 2001-09-26  
PRIORITY APPLICATION NUMBER: 60/323,580  
PRIORITY FILING DATE: 2001-09-19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 167  
LENGTH: 276  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-097-340-167

Query Match 100.0%; Score 1496; DB 15; Length 276;  
Best Local Similarity 100.0%; Pred. No. 5,7e-141;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAPHLHLSAAGBARALAKLPLMLQMAAEALPLPNDTRLDPEAYGAPCARSGOPMQ 60  
DB 1 MRAPHLHLSAAGBARALAKLPLMLQMAAEALPLPNDTRLDPEAYGAPCARSGOPMQ 60  
QY 61 VSLFNGSLFPCAGVLDQSVLTAACGKPKLWAVGDHLLLLQGEQLRTTRSVVHPK 120  
DB 61 VSLFNGSLFPCAGVLDQSVLTAACGKPKLWAVGDHLLLLQGEQLRTTRSVVHPK 120  
QY 121 YHSGSGILPRTRDEHMLKLARPVPGPRVRAQLPYRCAPDPDCCQVAGMGTTAAR 180  
DB 121 YHSGSGILPRTRDEHMLKLARPVPGPRVRAQLPYRCAPDPDCCQVAGMGTTAAR 180  
QY 181 RYKYNKGLTCSSTITLSPKCEVFYFGVVTNNMICAGLDKGDPCQSDSGPLVCDETIQ 240  
DB 181 RYKYNKGLTCSSTITLSPKCEVFYFGVVTNNMICAGLDKGDPCQSDSGPLVCDETIQ 240  
QY 241 GILSWGVPDGSAGHPAVYTQICKYMSWINKVIRSN 276  
DB 241 GILSWGVPDGSAGHPAVYTQICKYMSWINKVIRSN 276

RESULT 6  
US-10-021-368-11  
Sequence 11, Application US/10021368  
Publication No. US20020106367A1  
GENERAL INFORMATION:

APPLICANT: Band, Yimla  
TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
MOLECULES AND METHODS  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/021,368  
FILING DATE: 12-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/201,038  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00398/100002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 291 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 11:  
US-10-021-368-11

Query Match 96.2%; Score 1438.5; DB 14; Length 291;  
Best Local Similarity 94.8%; Pred. No. 3,3e-135;  
Matches 276; Conservative 0; Mismatches 0; Indels 15; Gaps 5;

QY 1 MRAPHLTSAASGABALAKLPLLMQMAAABALLPONTRLDPVAGS---PCARGSQ 57  
DB 1 MRAPHLTSAASGABALAKLPLLMQMAAABALLPONTRLDPVAGS---PCARGSQ 60  
QY 58 PMQVSLFNGSLFPCAGVLVDQSVLTAHCGNKPRLARVGDH---LLLLQSGQLRRTTR 114  
DB 61 PMQVSLFNGSLFPCAGVLVDQSVLTAHCGNKPRLARVGDH---LLLLQSGQLRRTTR 120  
QY 115 SVVHPKXHQSGGPIPRTEBHDMLKLKARVPV---GRRVALQLPFCAGPQGCQY 171  
DB 121 SVVHPKXHQSGGPIPRTEBHDMLKLKARVPV---GRRVALQLPFCAGPQGCQY 180  
QY 172 AGNGTARARVKNKGLTSSITILSPK---ECEVFYPGVYVNNMICAGLDRGQDPCQSD 228  
DB 181 AGNGTARARVKNKGLTSSITILSPKSECEVFYPGVYVNNMICAGLDRGQDPCQSD 240  
QY 229 SGGPLVDETLQGLILSGVT---PCGSAQHPAVYTOICKYMSWINKVIRSN 276  
DB 241 SGGPLVDETLQGLILSGVT---PCGSAQHPAVYTOICKYMSWINKVIRSN 291

RESULT 7  
US-09-946-374-194  
Sequence 194, Application US/09946374  
Publication No. US2003073129A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Baton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillen, Kenneth J.  
APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Matanabe, Colin K.  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2830P1C1  
CURRENT APPLICATION NUMBER: US/09/946,374  
CURRENT FILING DATE: 2001-09-04  
PRIOR APPLICATION NUMBER: 60/098716  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098723  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098749  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098750  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098803  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/098821  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/098843  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/099536  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099596  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099598  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099602  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099642  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099741  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099754  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099763  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099792  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099808  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099812  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099815  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099816  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/100385  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100388  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100390  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100584  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100627  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100661  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100662  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100664  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100683  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/100684  
PRIOR FILING DATE: 1998-09-17

```

PRIOR APPLICATION NUMBER: 60/100710
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100711
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100849
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100930
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101014
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101068
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101071
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101279
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/101471
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101472
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101474
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101475
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101476
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101479
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101743
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101915
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102207
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102240
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102307
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102330
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102331
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102484
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102487
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102570
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102571
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102684
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102687
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102965
PRIOR FILING DATE: 1998-10-02
PRIOR APPLICATION NUMBER: 60/103258
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103314
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103315

```

```

PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103328
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103395
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103396
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103401
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103449
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103633
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103678
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103679
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103711
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/104257
PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/104987
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105000
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105002
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105104
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807

```

Query Match 37.7%; Score 564; DB 11; Length 248;

Best Local Similarity 45.9%; Pred. No. 4.8e-48; Matches 118; Conservative 29; Mismatches 92; Indels 18; Gaps 5;

```

QY 23 LMAQLMAAEKALLPONDTRIDPEAY-GAPCAROSQPMOVSLEFNGSLFHCAGVLVDQSV 81
DB 7 LLLCVLGLSQA-----TPKIFNGTEGCRMSQPMOVLFGTSLRCGVLDIRBV 57
QY 82 LTAACGKPLMAKVGDDHLLQ-GEOLRTTSVHFKYHGGSEILRRDDEHML 140
DB 58 LTAACGSRMYVNLGHSLSQLDWTEQIRHSGFSVTHPYLGS-----TSHEDLRL 111
QY 141 LKLAPVVPGRVVALDLPYCAQPGDQCVAGWGTAAARVKYNKGLTCSITILSPKE 200
DB 112 LRLPLPRAVTSVCPPLPNDCAIAGTECHVSGGGINHPRNPFPLDQLNLSTISHAT 171
QY 201 CEVYFGVVTNNMIGLIDRGDPCSDSGSPLYCDBTLOGILSMG-VYRCGSGAHNAVY 259
DB 172 CHGVYPRGRTISNMVACAGVPGDQCGDSDGSPVLCGGVLDGLVMSGSGPGDGLPGVY 231
QY 260 TQICKYMSWINKVRSN 276
DB 232 TYICKYVDMIMMRNN 248

```

# RESULT 8

US-10-015-387A-194

; Sequence 194, Application US/10015387A

; Publication No. US20030135034A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

; APPLICANT: Eaton, Dan I.







PRIOR FILING DATE: 1998-10-22  
PRIOR APPLICATION NUMBER: 60/105266  
PRIOR FILING DATE: 1998-10-22  
PRIOR APPLICATION NUMBER: 60/105693  
PRIOR FILING DATE: 1998-10-26  
PRIOR APPLICATION NUMBER: 60/105694  
PRIOR FILING DATE: 1998-10-26  
PRIOR APPLICATION NUMBER: 60/105807  
PRIOR FILING DATE: 1998-10-27  
PRIOR APPLICATION NUMBER: 60/105861  
PRIOR FILING DATE: 1998-10-27  
PRIOR APPLICATION NUMBER: 60/105882  
PRIOR FILING DATE: 1998-10-27  
PRIOR APPLICATION NUMBER: 60/106023  
PRIOR FILING DATE: 1998-10-28

Query Match 37.7%; Score 564; DB 12; Length 248;  
Best Local Similarity 45.9%; Pred. No. 4,8e-48;  
Matches 118; Conservative 29; Mismatches 92; Indels 18; Gaps 5;

QY 23 LMAQLMAEALALPQNDTRLDPEAY-GAPCARSGSPQVQVSLFNGLSFHCAGVLVDQSWV 81  
DB 7 LLLCVLGLSQA-----TPKIFNGTECGRNSQPMQVGLFEGTSLRCGGLVDHKNV 57  
QY 82 LTAHCGNKPLMARVGDHLLLO-GEOLKRTTSVVAHKYHQSGLPILPRRTDEHDL 140  
DB 58 LTAHCGSRVWVLGEHSLSQLDWTQIRHSGSVTHPGYLGAS-----TSHHDLRL 111  
QY 141 LKLAPVYVGRVRLALPYRCAGPQDQCVAGNMTAARKYKNGJLTCSITLISPK 200  
DB 112 LRLRPVRYTSVQPLPNCATAGTECHVSGWGITNHPNPFDLLQCLNLSIVSHAT 171  
QY 201 CEVFPYGVVTTNNMTCAGLDRCQDPCQSDGSPVLCDETLQGLISWG-VYPCGSAQHPAVY 259  
DB 172 CHGVYFGRITSNMVCAGVPGQDAQCGSDGSPVLCVGLQGLVSWGSGVPCGQDGIPIGY 231  
QY 260 TOICKYMSINKVIRSN 276  
DB 232 TYICKYVDWIMIRMRN 248

RESULT 11  
US-10-015-392A-194  
Sequence 194, Application US/10015392A  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Baton, Dan I.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: P2830P1C58  
CURRENT APPLICATION NUMBER: US/10/015,392A  
CURRENT FILING DATE: 2001-12-12  
PRIOR APPLICATION NUMBER: 60/098716  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098723  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098749  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098750  
PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098803  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/098821  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/098843  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/099536  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099596  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099598  
PRIOR FILING DATE: 1998-09-09  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 477  
SEQ ID NO 194  
LENGTH: 248  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-015-392A-194

Query Match 37.7%; Score 564; DB 12; Length 248;  
Best Local Similarity 45.9%; Pred. No. 4,8e-48;  
Matches 118; Conservative 29; Mismatches 92; Indels 18; Gaps 5;

QY 23 LMAQLMAEALALPQNDTRLDPEAY-GAPCARSGSPQVQVSLFNGLSFHCAGVLVDQSWV 81  
DB 7 LLLCVLGLSQA-----TPKIFNGTECGRNSQPMQVGLFEGTSLRCGGLVDHKNV 57  
QY 82 LTAHCGNKPLMARVGDHLLLO-GEOLKRTTSVVAHKYHQSGLPILPRRTDEHDL 140  
DB 58 LTAHCGSRVWVLGEHSLSQLDWTQIRHSGSVTHPGYLGAS-----TSHHDLRL 111  
QY 141 LKLAPVYVGRVRLALPYRCAGPQDQCVAGNMTAARKYKNGJLTCSITLISPK 200  
DB 112 LRLRPVRYTSVQPLPNCATAGTECHVSGWGITNHPNPFDLLQCLNLSIVSHAT 171  
QY 201 CEVFPYGVVTTNNMTCAGLDRCQDPCQSDGSPVLCDETLQGLISWG-VYPCGSAQHPAVY 259  
DB 172 CHGVYFGRITSNMVCAGVPGQDAQCGSDGSPVLCVGLQGLVSWGSGVPCGQDGIPIGY 231  
QY 260 TOICKYMSINKVIRSN 276  
DB 232 TYICKYVDWIMIRMRN 248

RESULT 12  
US-10-017-253A-194  
Sequence 194, Application US/10017253A  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Baton, Dan I.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: P2830P1C62  
CURRENT APPLICATION NUMBER: US/10/017,253A  
CURRENT FILING DATE: 2001-12-13  
PRIOR APPLICATION NUMBER: 60/098716  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098723  
PRIOR FILING DATE: 1998-09-01

```

1      RESULT 14
2      US-10-012-064A-194
3      ; Sequence 194, Application US/10012064A
4      ; Publication No. US20030180836A1
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Baker, Kevin P.
7      ; APPLICANT: Botstein, David
8      ; APPLICANT: Desnuyers, Luc
9      ; APPLICANT: Eaton, Dan I.
10     ; APPLICANT: Ferrara, Napoleone
11     ; APPLICANT: Fong, Sherman
12     ; APPLICANT: Gao, Wei-Qiang
13     ; APPLICANT: Goddard, Audrey
14     ; APPLICANT: Godowski, Paul J.
15     ; APPLICANT: Grimaldi, Christopher J.
16     ; APPLICANT: Gunney, Austin B.
17     ; APPLICANT: Hillan, Kenneth J.
18     ; APPLICANT: Pan, James
19     ; APPLICANT: Paoni, Nicholas F.
20     ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
21     ; FILE OF INVENTION: Acids Encoding the Same
22     ; FILE REFERENCE: P2830P1C19
23     ; CURRENT APPLICATION NUMBER: US/10/012,064A
24     ; CURRENT FILING DATE: 2002-07-15
25     ; PRIOR APPLICATION NUMBER: 60/098716
26     ; PRIOR FILING DATE: 1998-09-01
27     ; PRIOR APPLICATION NUMBER: 60/098723
28     ; PRIOR FILING DATE: 1998-09-01
29     ; PRIOR APPLICATION NUMBER: 60/098749
30     ; PRIOR FILING DATE: 1998-09-01
31     ; PRIOR APPLICATION NUMBER: 60/098750
32     ; PRIOR FILING DATE: 1998-09-01
33     ; PRIOR APPLICATION NUMBER: 60/098803
34     ; PRIOR FILING DATE: 1998-09-02
35     ; PRIOR APPLICATION NUMBER: 60/098821
36     ; PRIOR FILING DATE: 1998-09-02
37     ; PRIOR APPLICATION NUMBER: 60/098843
38     ; PRIOR FILING DATE: 1998-09-02
39     ; PRIOR APPLICATION NUMBER: 60/099536
40     ; PRIOR FILING DATE: 1998-09-09

```

PRIOR APPLICATION NUMBER: 60/099596  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099598  
PRIOR FILING DATE: 1998-09-09  
Prior Application data removed - See File Wrapper or PAM.  
NUMBER OF SEQ ID NOS: 477  
SEQ ID NO 194  
LENGTH: 248  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-012-064A-194

Query Match 37.7%; Score 564; DB 12; Length 248;  
Best Local Similarity 45.9%; Pred. No. 4,8e-48;

Matches 118; Conservative 29; Mismatches 92; Indels 18; Gaps 5;

QY 23 LMAQMAEAPALLPQNDTRLOPEAY-GAPCARSGQPMQVSLFNGLSFRCAGVLVDQSY 81  
DB 7 LLLCVLGLSQAA-----TPKIFNGTECGSNQPMQVGLFEGTSLRCGGLIDHRMV 57  
QY 82 LTTAHCGRKPLMARVGDHLLIQ-GEQLRTTRSVVHKYQSGPILPRITDHDML 140  
DB 58 LTTAHCGRKPLMARVGDHLLIQ-GEQLRTTRSVVHKYQSGPILPRITDHDML 140  
QY 141 LKTLARPVDPVRALQLPYCAQPDQCVAGMTTAAARKYKXKGLTCSITILSPKE 200  
DB 112 LRLRLPYRTSSVQPLPLPDCATAGTBECHVSGMTNHPNPDLGLNLISVSHAT 171  
QY 201 CEFFYPVNTNNMTICAGLDGDPCCSDSGGPLYCDETIQGLISNG-VTPCGSAHPANY 259  
DB 172 CHGVYPRITSSNWCAGVGGVGDACQDSGGPLVCGVLQGLVSWGSVAPCGDGPVGY 231  
QY 260 TQICKYMSWINKYRSN 276  
DB 232 TYICKYVDIMIRMRN 248

RESULT 15  
US-10-017-867A-194

Sequence 194, Application US/10017867A  
Publication No. US20030180792A1  
GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan I.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Goddard, Audrey  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Guiney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
Acids Encoding the Same  
FILE REFERENCE: P2830P1C60  
CURRENT APPLICATION NUMBER: US/10/017,867A  
PRIOR APPLICATION NUMBER: 60/098716  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098723  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098749  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098750  
PRIOR FILING DATE: 1998-09-01  
PRIOR APPLICATION NUMBER: 60/098803  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/098821  
PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/098843  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/099536  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099596  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099598  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099602  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099642  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099741  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099754  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099763  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099792  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099808  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099812  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099815  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099816  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/100385  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100388  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100390  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100584  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100627  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100661  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100662  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100664  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100683  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/100684  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/100710  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/100711  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/100848  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/100849  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/100919  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/100930  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/101014  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/101068  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/101071  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/101279  
PRIOR FILING DATE: 1998-09-22  
PRIOR APPLICATION NUMBER: 60/101471  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101472  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101474

```

; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101475
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101476
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101477
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101479
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101741
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101743
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101915
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/102207
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102240
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102307
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102330
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102331
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102484
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102487
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102570
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102571
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 60/102684
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/102687
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/102965
; PRIOR FILING DATE: 1998-10-02
; PRIOR APPLICATION NUMBER: 60/103258
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103314
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103315
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103328
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103395
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103396
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103401
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 60/103449
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/103633
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103678
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103679
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/103711
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/104257
; PRIOR FILING DATE: 1998-10-14
; PRIOR APPLICATION NUMBER: 60/104987
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105000
; PRIOR FILING DATE: 1998-10-20

```

```

; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105104
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 60/105169
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105266
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105693
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105694
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105807
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: 60/105881
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: 60/105882
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: 60/106023
; PRIOR FILING DATE: 1998-10-28

```

```

Query Match 37.7%; Score 564; DB 12; Length 248;
Best Local Similarity 45.9%; Pred. No. 4, 8e-48;
Matches 118; Conservative 29; Mismatches 92; Indels 16; Gaps 5;

```

```

QY 23 LMAQLMAAARALLPQNDTRLDPEAY-GAPCARGSQPMOVSLENGLSFHCAGLVNDQSV 81
DB 7 LILCVLGLSQA-----TKIFNGTECGNSQPMQVGLDEGSLRCGSLIDHRV 57
QY 82 LTAHAGNKPMLWAVGDHLLLIQ-GEOLRTTRSVVHPKXHGSGPILPRTDEHML 140
DB 58 LTAHCSGRVYVRLBHSLSQLDWTGEQIRHSGFSVTHPGYLAS-----TSHEDRL 111
QY 141 LKARPYVGPFRRAQLPYRCAGQPDQCVAGMGTAAARVYKNGLTCSSITILSPKE 200
DB 112 LRLPLRVYTSVQPLPLEPDCATAGTECHVSQMGITNHRNPPDLLQCLNLSIVSHAT 171
QY 201 CEVFYEGVVTNNMICAGLDRGQDPCOSDSGGLVCEETLQGLISWG-VYECGSAQHFAVY 259
DB 172 CHEVYFERITSNNVACGAGVVGQDACCQDSGGLVCGSLVGLVSKSVGRCQDGLPGVY 231
QY 260 TQICKTMSWINKVIRSN 276
DB 232 TYICKYVDMIRMIMNN 248

```

```

Search completed: November 25, 2003, 09:13:57
Job time : 48 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 09:06:20 / Search time 22:Seconds

(without alignments)  
530.809 Million cell updates/sec

Title: US-10-021-368-1

Perfect score: 1496

Sequence: 1 MRAPHLHLSAASGARALAKL.....AVYTQICKYMSWINKVIREN 276

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A-COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B-COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5A-COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/5B-COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS-COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfiltest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1496	100.0	276	1	US-08-467-155A-1
2	1496	100.0	276	2	US-08-628-198-1
3	1496	100.0	276	3	US-09-301-038-1
4	1496	100.0	276	3	PCT-US96-07345-1
5	1438.5	96.2	291	1	US-08-467-155A-11
6	1438.5	96.2	291	2	US-08-628-198-11
7	1438.5	96.2	291	3	US-09-201-038-11
8	1438.5	96.2	291	5	PCT-US96-07345-11
9	549	36.7	253	2	US-08-557-146-2
10	549	36.7	253	2	US-08-824-874-3
11	549	36.7	253	2	US-09-154-344-2
12	549	36.7	253	3	US-08-330-188-2
13	549	36.7	253	3	US-09-210-084-3
14	549	36.7	253	4	US-09-764-762-3
15	549	36.7	253	5	PCT-US96-04294-2
16	539	36.0	260	3	US-09-008-271A-7
17	538	36.0	260	3	US-09-070-526-2
18	538	36.0	288	4	US-09-386-642-13
19	533	35.6	224	4	US-08-944-483-33
20	533	35.6	225	2	US-09-027-337-4
21	533	35.6	225	4	US-09-644-600-4
22	522.5	34.9	225	2	US-08-557-146-12
23	522.5	34.9	225	2	US-09-154-344-12
24	513	34.3	260	3	US-09-025-059-3
25	507.5	33.9	282	4	US-09-025-059-1
26	503.5	33.7	250	4	US-09-205-258-427
27	500	33.4	289	4	US-09-386-642-14

## ALIGNMENTS

28	493	33.0	247	2	US-08-956-267A-2	Sequence 2, App11
29	491	32.8	223	1	US-08-278-091-9	Sequence 9, App11
30	491	32.8	223	1	US-08-483-855-9	Sequence 9, App11
31	491	32.8	223	1	US-08-472-173-9	Sequence 9, App11
32	491	32.8	223	2	US-08-487-167-9	Sequence 9, App11
33	491	32.8	223	2	US-08-482-816-9	Sequence 9, App11
34	491	32.8	223	2	US-08-296-149-9	Sequence 9, App11
35	491	32.8	223	2	US-08-801-499-9	Sequence 9, App11
36	491	32.8	223	2	US-08-615-271-9	Sequence 9, App11
37	491	32.8	223	3	US-09-074-660-9	Sequence 9, App11
38	491	32.8	223	3	US-09-074-659-9	Sequence 9, App11
39	491	32.8	223	3	US-09-106-468-9	Sequence 9, App11
40	491	32.8	223	3	US-09-106-466A-9	Sequence 9, App11
41	491	32.8	223	3	US-09-106-467-9	Sequence 9, App11
42	491	32.8	223	3	US-09-120-582-2	Sequence 2, App11
43	488.5	32.7	248	3	US-08-944-483-24	Sequence 24, App11
44	488.5	32.7	293	4	US-09-996-243-309	Sequence 309, App11
45	482	32.2	246	2	US-08-978-404B-44	Sequence 44, App11

```
RESULT 1
US-08-467-155A-1
; Sequence 1, Application US/08467155A
; Patent No. 5736377
; GENERAL INFORMATION:
; APPLICANT: Band, Viola
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED
; TITLES OF INVENTION: MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,155A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00398/100001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEO ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 276 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-467-155A-1

Query Match      100.0%; Score 1496; DB 1; Length 276;
Best Local Similarity 100.0%; Pred. No. 3.4e-146;
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MRAPHLHLSAASGARALAKLPLMAQDMAEALPLPNDTRLPDEAYGACARSSQWQ 60
DB      1 MRAPHLHLSAASGARALAKLPLMAQDMAEALPLPNDTRLPDEAYGACARSSQWQ 60
```

QY 61 VSLFNGLSFHCAGVLDQSWVLTAAHCGNKPIMARVGDHLLILQSGQLRRTTRSVVHPK 120  
DB 61 VSLFNGLSFHCAGVLDQSWVLTAAHCGNKPIMARVGDHLLILQSGQLRRTTRSVVHPK 120  
QY 121 YHSGSGPILPRRTDEHMLLKLARPVVPGPRVALQLPYRCAGPDCCQVAGGTTAAR 180  
DB 121 YHSGSGPILPRRTDEHMLLKLARPVVPGPRVALQLPYRCAGPDCCQVAGGTTAAR 180  
QY 181 RVKYNKGLTGSITLISPEKCEVFYPGVVTNNMICAGLDRGDDPCQSDSGGLVCDETLQ 240  
DB 181 RVKYNKGLTGSITLISPEKCEVFYPGVVTNNMICAGLDRGDDPCQSDSGGLVCDETLQ 240  
QY 241 GILSWGVPQCSAQHPAVYTOICKYMSWINKYIRSN 276  
DB 241 GILSWGVPQCSAQHPAVYTOICKYMSWINKYIRSN 276

## RESULT 2

US-08-628-198-1  
Sequence 1, Application US/08628198  
Patent No. 5843694  
GENERAL INFORMATION:  
APPLICANT: Band, Vimala  
TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
TITLE OF INVENTION: MOLECULES AND METHODS  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/628.198  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/667,155  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00398/100002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 276 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-628-198-1

Query Match 100.0%; Score 1496; DB 2; Length 276;  
Best Local Similarity 100.0%; Pred. No. 3.4e-146;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAAPHHLISAASGARALAKLPLMAQLMAAEALLPQNDTRLDPDEAYGAPCARSGOPWQ 60  
DB 1 MRAAPHHLISAASGARALAKLPLMAQLMAAEALLPQNDTRLDPDEAYGAPCARSGOPWQ 60  
QY 61 VSLFNGLSFHCAGVLDQSWVLTAAHCGNKPIMARVGDHLLILQSGQLRRTTRSVVHPK 120  
DB 61 VSLFNGLSFHCAGVLDQSWVLTAAHCGNKPIMARVGDHLLILQSGQLRRTTRSVVHPK 120

QY 121 YHSGSGPILPRRTDEHMLLKLARPVVPGPRVALQLPYRCAGPDCCQVAGGTTAAR 180  
DB 121 YHSGSGPILPRRTDEHMLLKLARPVVPGPRVALQLPYRCAGPDCCQVAGGTTAAR 180  
QY 181 RVKYNKGLTGSITLISPEKCEVFYPGVVTNNMICAGLDRGDDPCQSDSGGLVCDETLQ 240  
DB 181 RVKYNKGLTGSITLISPEKCEVFYPGVVTNNMICAGLDRGDDPCQSDSGGLVCDETLQ 240  
QY 241 GILSWGVPQCSAQHPAVYTOICKYMSWINKYIRSN 276  
DB 241 GILSWGVPQCSAQHPAVYTOICKYMSWINKYIRSN 276

## RESULT 3

US-09-201-038-1  
Sequence 1, Application US/09201038  
Patent No. 615387  
GENERAL INFORMATION:  
APPLICANT: Band, Vimala  
TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
TITLE OF INVENTION: MOLECULES AND METHODS  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/201.038  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/628,198  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00398/100002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 276 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-201-038-1

Query Match 100.0%; Score 1496; DB 3; Length 276;  
Best Local Similarity 100.0%; Pred. No. 3.4e-146;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAAPHHLISAASGARALAKLPLMAQLMAAEALLPQNDTRLDPDEAYGAPCARSGOPWQ 60  
DB 1 MRAAPHHLISAASGARALAKLPLMAQLMAAEALLPQNDTRLDPDEAYGAPCARSGOPWQ 60  
QY 61 VSLFNGLSFHCAGVLDQSWVLTAAHCGNKPIMARVGDHLLILQSGQLRRTTRSVVHPK 120  
DB 61 VSLFNGLSFHCAGVLDQSWVLTAAHCGNKPIMARVGDHLLILQSGQLRRTTRSVVHPK 120  
QY 121 YHSGSGPILPRRTDEHMLLKLARPVVPGPRVALQLPYRCAGPDCCQVAGGTTAAR 180  
DB 121 YHSGSGPILPRRTDEHMLLKLARPVVPGPRVALQLPYRCAGPDCCQVAGGTTAAR 180

Db 121 YHOGSGFLBRTDEHMLKLARPVPGPRVRLQLPYRCAQPGDQCQVAGWGTAAAR 180  
QY 181 RVKYNKGLTSSITTLSPKCEVFYRGVVTNNMICAGLDRGDPCCSDSGPLVCDETLQ 240  
Db 181 RVKYNKGLTSSITTLSPKCEVFYRGVVTNNMICAGLDRGDPCCSDSGPLVCDETLQ 240  
QY 241 GILSWGVPCCSAQHPAVYTQICKYMSWINKVIRSN 276  
Db 241 GILSWGVPCCSAQHPAVYTQICKYMSWINKVIRSN 276

RESULT 4  
PCT-US96-07343-1

/ Sequence 1, Application PC/US9607343  
/ GENERAL INFORMATION:  
/ APPLICANT: New England Medical Center Hospitals, Inc.  
/ TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
/ NUMBER OF SEQUENCES: 11  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Fish & Richardson P.C.  
/ STREET: 225 Franklin Street  
/ CITY: Boston  
/ STATE: MA  
/ COUNTRY: USA  
/ ZIP: 02110-2804  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: PCT/US96/07343  
/ FILING DATE:  
/ CLASSIFICATION:  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/467,155  
/ FILING DATE: 06-JUN-1995  
/ CLASSIFICATION:  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Clark, Paul T.  
/ REGISTRATION NUMBER: 30,162  
/ REFERENCE/DOCKET NUMBER: 00398/100001  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 617/542-5070  
/ TELEFAX: 617/542-8906  
/ TELEX: 200154  
/ INFORMATION FOR SEQ ID NO: 1:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 276 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: not relevant  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: protein  
/ PCT-US96-07343-1

Query Match 100.0%; Score 1496; DB 5; Length 276;  
Best Local Similarity 100.0%; Pred. No. 3.4e-146;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAPHLHLSAASGABALAKLPLMAQMAAFAALLPONDRLDPEAYGAPCARGSQPMQ 60  
Db 1 MRAPHLHLSAASGABALAKLPLMAQMAAFAALLPONDRLDPEAYGAPCARGSQPMQ 60  
QY 61 VSLFNGLSFHCAQGLVDQSWVLTAAHCGNKPIMARVGDHLLLLQSEQLRRTTRSVVHPK 120  
Db 61 VSLFNGLSFHCAQGLVDQSWVLTAAHCGNKPIMARVGDHLLLLQSEQLRRTTRSVVHPK 120  
QY 121 YHOGSGFLBRTDEHMLKLARPVPGPRVRLQLPYRCAQPGDQCQVAGWGTAAAR 180  
Db 121 YHOGSGFLBRTDEHMLKLARPVPGPRVRLQLPYRCAQPGDQCQVAGWGTAAAR 180  
QY 181 RVKYNKGLTSSITTLSPKCEVFYRGVVTNNMICAGLDRGDPCCSDSGPLVCDETLQ 240

Db 181 RVKYNKGLTSSITTLSPKCEVFYRGVVTNNMICAGLDRGDPCCSDSGPLVCDETLQ 240  
QY 241 GILSWGVPCCSAQHPAVYTQICKYMSWINKVIRSN 276  
Db 241 GILSWGVPCCSAQHPAVYTQICKYMSWINKVIRSN 276

RESULT 5  
US-08-467-155A-11

/ Sequence 11, Application US/08467155A  
/ Patent No. 5736377  
/ GENERAL INFORMATION:  
/ APPLICANT: Band, Viola  
/ TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
/ NUMBER OF SEQUENCES: 11  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Fish & Richardson P.C.  
/ STREET: 225 Franklin Street  
/ CITY: Boston  
/ STATE: MA  
/ COUNTRY: USA  
/ ZIP: 02110-2804  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/467,155A  
/ FILING DATE: 06-JUN-1995  
/ CLASSIFICATION: 435  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Clark, Paul T.  
/ REGISTRATION NUMBER: 30,162  
/ REFERENCE/DOCKET NUMBER: 00398/100001  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 617/542-5070  
/ TELEFAX: 617/542-8906  
/ TELEX: 200154  
/ INFORMATION FOR SEQ ID NO: 11:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 291 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: not relevant  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: protein  
/ US-08-467-155A-11

Query Match 96.2%; Score 1438.5; DB 1; Length 291;  
Best Local Similarity 94.8%; Pred. No. 3.1e-140;  
Matches 276; Conservative 0; Mismatches 0; Indels 15; Gaps 5;

QY 1 MRAPHLHLSAASGABALAKLPLMAQMAAFAALLPONDRLDPEAYGAPCARGSQ 57  
Db 1 MRAPHLHLSAASGABALAKLPLMAQMAAFAALLPONDRLDPEAYGAPCARGSQ 60  
QY 58 PMOVSLENGLSFHCAQGLVDQSWVLTAAHCGNKPIMARVGDHLLLLQSEQLRRTTR 114  
Db 61 PMOVSLENGLSFHCAQGLVDQSWVLTAAHCGNKPIMARVGDHLLLLQSEQLRRTTR 120  
QY 115 SVYHKKYHOGSGFLBRTDEHMLKLARPVPGPRVRLQLPYRCAQPGDQCQV 171  
Db 121 SVYHKKYHOGSGFLBRTDEHMLKLARPVPGPRVRLQLPYRCAQPGDQCQV 180  
QY 172 AGWGTAAARVYKYNKGLTSSITTLSPKCEVFYRGVVTNNMICAGLDRGDPCCSD 228  
Db 181 AGWGTAAARVYKYNKGLTSSITTLSPKCEVFYRGVVTNNMICAGLDRGDPCCSD 240  
QY 229 SGGPLVCDETLQILSWGVPCCSAQHPAVYTQICKYMSWINKVIRSN 276  
Db 241 SGGPLVCDETLQILSWGVPCCSAQHPAVYTQICKYMSWINKVIRSN 291

```
RESULT 6
US-08-628-198-11
; Sequence 11, Application US/08628198
; Patent No. 5843694
; GENERAL INFORMATION:
; APPLICANT: Band, Vimala
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED
; MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,198
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/467,155
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00398/100002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-628-198-11

Query Match 96.2%; Score 1438.5; DB 2; Length 291;
Best Local Similarity 94.8%; Pred. No. 3.1e-140;
Matches 276; Conservative 0; Mismatches 0; Indels 15; Gaps 5;

QY 1 MRAPIHLISAAGARALAKLLPLMAQIWAABALLPQNDTRLDPEAYGA--PCARGSQ 57
DB 1 MRAPIHLISAAGARALAKLLPLMAQIWAABALLPQNDTRLDPEAYGANESPCARGSQ 60
QY 58 PMQVSLFNGISFPCAGVLDQSWVLTAAHCGNKPLMARVGDH---LILLQSGQLRRTTR 114
DB 61 PMQVSLFNGISFPCAGVLDQSWVLTAAHCGNKPLMARVGDHESILLQSGQLRRTTR 120
QY 115 SVVHPKXHQSGPILPRRTDEHDLMLKLARPVV--GPRVALQLPYRCAQPGDCCQV 171
DB 121 SVVHPKXHQSGPILPRRTDEHDLMLKLARPVVFNESGPRVALQLPYRCAQPGDCCQV 180
QY 172 AGNGTTAARVKNKGLTCSITILSPK--ECVEVYFGVVTNNMICAGLDRODPCQSD 228
DB 181 AGNGTTAARVKNKGLTCSITILSPKNESECEVFYGVVTNNMICAGLDRODPCQSD 240
QY 229 SGGPLVCDETLQGISLNGVY--PCGSAQHPAVYTOICKYMSWINKYIRSN 276
DB 241 SGGPLVCDETLQGISLNGVYNESECEVFYGVVTTOICKYMSWINKYIRSN 291

RESULT 7
```

```
US-09-201-038-11
; Sequence 11, Application US/09201038
; Patent No. 6153387
; GENERAL INFORMATION:
; APPLICANT: Band, Vimala
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED
; MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,038
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/628,198
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00398/100002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-201-038-11

Query Match 96.2%; Score 1438.5; DB 3; Length 291;
Best Local Similarity 94.8%; Pred. No. 3.1e-140;
Matches 276; Conservative 0; Mismatches 0; Indels 15; Gaps 5;

QY 1 MRAPIHLISAAGARALAKLLPLMAQIWAABALLPQNDTRLDPEAYGA--PCARGSQ 57
DB 1 MRAPIHLISAAGARALAKLLPLMAQIWAABALLPQNDTRLDPEAYGANESPCARGSQ 60
QY 58 PMQVSLFNGISFPCAGVLDQSWVLTAAHCGNKPLMARVGDH---LILLQSGQLRRTTR 114
DB 61 PMQVSLFNGISFPCAGVLDQSWVLTAAHCGNKPLMARVGDHESILLQSGQLRRTTR 120
QY 115 SVVHPKXHQSGPILPRRTDEHDLMLKLARPVV--GPRVALQLPYRCAQPGDCCQV 171
DB 121 SVVHPKXHQSGPILPRRTDEHDLMLKLARPVVFNESGPRVALQLPYRCAQPGDCCQV 180
QY 172 AGNGTTAARVKNKGLTCSITILSPK--ECVEVYFGVVTNNMICAGLDRODPCQSD 228
DB 181 AGNGTTAARVKNKGLTCSITILSPKNESECEVFYGVVTNNMICAGLDRODPCQSD 240
QY 229 SGGPLVCDETLQGISLNGVY--PCGSAQHPAVYTOICKYMSWINKYIRSN 276
DB 241 SGGPLVCDETLQGISLNGVYNESECEVFYGVVTTOICKYMSWINKYIRSN 291

RESULT 8
PCT-US96-07343-11
; Sequence 11, Application PC/TUS9607343
; GENERAL INFORMATION:
```



APPLICANT: New England Medical Center Hospitals, Inc.  
TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/07343  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/467,155  
FILING DATE: 06-JUN-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00398/100001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-8070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 291 amino acids  
TYPE: amino acid  
STANDARDS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-07343-11

Query Match 96.2%; Score 1438.5; DB 5; Length 291;  
Best Local Similarity 94.8%; Pred. No. 3.1e-140;  
Matches 276; Conservative 0; Mismatches 0; Indels 15; Gaps 5;

QY 1 MRAPHLHSASAGARALAKLPLMLAQLMAEALPLPNDTRLDPBAYGA---PCARGSQ 57  
DB 1 MRAPHLHSASAGARALAKLPLMLAQLMAEALPLPNDTRLDPBAYGANESPCARGSQ 60  
QY 58 PMQVSLFNGLSFHCAGVLVDQSNVLTAAHCNKPLMARVGDH---LILLQGEQDRLTR 114  
DB 61 PMQVSLFNGLSFHCAGVLVDQSNVLTAAHCNKPLMARVGDHNSILLQGEQDRLTR 120  
QY 115 SVVHPKYHGGSPILPRTRDEHDMILKLARVPV---GPRVRAIOLPRCAQDPDQCY 171  
DB 121 SVVHPKYHGGSPILPRTRDEHDMILKLARVPVPESSPRVRAIOLPRCAQDPDQCY 180  
QY 172 AGMGTTAARVYKXNKGLTCSITILSPK---ECVFPYPGVVTNNMICAGLDGQDPCQSD 228  
DB 181 AGMGTTAARVYKXNKGLTCSITILSPKSECEVFYPPVYNNMICAGLDGQDPCQSD 240  
QY 229 SGGPLVCDLTLQGLISMGVY---PCGSAQHPAVYTIQICKYMWINKVIRSN 276  
DB 241 SGGPLVCDLTLQGLISMGVYNESPCGSAQHPAVYTIQICKYMWINKVIRSN 291

RESULT 9  
US-08-557-146-2  
Sequence 2, Application US/08557146  
Patent No. 5834290  
GENERAL INFORMATION:  
APPLICANT: Egeltud, Torbjorn  
APPLICANT: Hansson, Lennart

TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic  
TITLE OF INVENTION: Enzyme (SCCE)  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sterner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-557-146-2

Query Match 36.7%; Score 549; DB 2; Length 253;  
Best Local Similarity 40.9%; Pred. No. 1.2e-48;  
Matches 108; Conservative 45; Mismatches 97; Indels 14; Gaps 4;

QY 14 ARALAKLPLMAQMAEALPLPNDTRLDPBAYGACARGSQPMQVSLFNGLSFHCAG 73  
DB 2 ARSLPLPQILILSLALETFAGBRAQGDKITD---GACARGSHPMQVALISGNOLHCGG 57  
QY 74 VLVQSWVLTAHAGCNKELMARVGDHLLILLQGEQDRLTRSVHPKYHGGSPILPRRT 133  
DB 58 VLVNERVLTAAHCQNEVYTHLGSDDLGDRAQRI-KASSFRHPG-----STQT 108  
QY 134 DEHDMILKLARVPVPGRRVRAIOLPRCAQDPDQCYAGMGTTAARVYKXNKGLTCSGI 193  
DB 109 HNDMLVYKXNSQARLSMVYKVRLPSCRCEPFGITCTVSGMGTTSPVTFPSDLMCVDV 168  
QY 194 TILSPKECEVYPGVVTNNMICAGL-DRGQDPCQSDSGGPLVCDLTLQGLISMGVYPCGS 252  
DB 169 KLISPDQCTKYKDLLENLSMLCAGIPDSKKAACGSDGPLVCHGTLQGLVSMGTFPCGQ 228  
QY 253 AQHPAVYTIQICKYMWINKVIRSN 276  
DB 229 PNDPGVYTVCKCFKWKINDTWKKH 252

RESULT 10  
US-08-824-874-3  
Sequence 3, Application US/08824874  
Patent No. 5962300  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: NOVEL KALLIKREIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA

COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/824,874  
FILING DATE: Filed Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
US-08-824-874-3

Query Match 36.7%; Score 549; DB 2; Length 253;  
Best Local Similarity 40.9%; Pred. No. 1.2e-48;  
Matches 108; Conservative 45; Mismatches 97; Indels 14; Gaps 4;

QY 14 ARAALAKLPLMAQLMAAERALLPQNDRLDPEAYGAPCARSGQPMQVSLFNGLSFHCAG 73  
DB 2 ARSLPLPLQILILSLALETAGEBAQSDKIID---GAPCARSGHPMQVALISNGQLHCAG 57  
QY 74 VLVDQSVLTLAAHCGNKPMAVGVDDHLLLGEOQLRRTTRSVVHPYHQSGPILPRT 133  
DB 58 VLAVNRVWLTLAAHCKNMEYTVHLGSDTLGDRRAQRI-KASKSPRHFGY-----STQT 108  
QY 134 DEHDLMLKLARPVVGPVRVRLQLPYRCAQPGDQCQVAGMTTAAARKYKNGLTCSST 193  
DB 109 HVDNMLVTLNLSQARLSNVKVRLPSCRPEPTTIVSGMGTTSPTDVFPSDLNCVDV 168  
QY 194 TILSPKECEVFYFGVATNNMTCAGL-DRGQDPQSDSGGPLYCDETLQGLISGVYPCGS 252  
DB 169 KLISPDCTKVKYKDLLENMTCAGIPDSKKNACNGDSGGPLYCRGTLQGLVSWGTFCGQ 228  
QY 253 AQHPAVYTOICKYMSWINKVIRSN 276  
DB 229 PNDGVTYQVCKFTKWINDTMKKH 252

RESULT 11  
US-09-154-344-2  
Sequence 2, Application US/09154344  
Patent No. 5981256  
GENERAL INFORMATION:  
APPLICANT: Egeltud, Torbjorn  
APPLICANT: Hansson, Lennart  
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.

ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/154,344  
FILING DATE: 16-SEP-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Steiner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-154-344-2

Query Match 36.7%; Score 549; DB 2; Length 253;  
Best Local Similarity 40.9%; Pred. No. 1.2e-48;  
Matches 108; Conservative 45; Mismatches 97; Indels 14; Gaps 4;

QY 14 ARAALAKLPLMAQLMAAERALLPQNDRLDPEAYGAPCARSGQPMQVSLFNGLSFHCAG 73  
DB 2 ARSLPLPLQILILSLALETAGEBAQSDKIID---GAPCARSGHPMQVALISNGQLHCAG 57  
QY 74 VLVDQSVLTLAAHCGNKPMAVGVDDHLLLGEOQLRRTTRSVVHPYHQSGPILPRT 133  
DB 58 VLAVNRVWLTLAAHCKNMEYTVHLGSDTLGDRRAQRI-KASKSPRHFGY-----STQT 108  
QY 134 DEHDLMLKLARPVVGPVRVRLQLPYRCAQPGDQCQVAGMTTAAARKYKNGLTCSST 193  
DB 109 HVDNMLVTLNLSQARLSNVKVRLPSCRPEPTTIVSGMGTTSPTDVFPSDLNCVDV 168  
QY 194 TILSPKECEVFYFGVATNNMTCAGL-DRGQDPQSDSGGPLYCDETLQGLISGVYPCGS 252  
DB 169 KLISPDCTKVKYKDLLENMTCAGIPDSKKNACNGDSGGPLYCRGTLQGLVSWGTFCGQ 228  
QY 253 AQHPAVYTOICKYMSWINKVIRSN 276  
DB 229 PNDGVTYQVCKFTKWINDTMKKH 252

RESULT 12  
US-08-930-188-2  
Sequence 2, Application US/08930188  
Patent No. 6093397  
GENERAL INFORMATION:  
APPLICANT: Dixon, Eric P.  
APPLICANT: Johnstone, Edward M.  
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND  
RELATED NUCLEIC ACIDS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/930,188
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,257
FILING DATE: 04-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Bialock, Donna K.
REGISTRATION NUMBER: 38,082
REFERENCE/DOCKET NUMBER: X9239
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-277-1090
TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-930-188-2

Query Match
Best Local Similarity 36.7%; Score 549; DB 3; Length 253;
Matches 108; Conservative 45; Mismatches 97; Indels 14; Gaps 4;

QY 14 ARALAKLPLMAQMAEAALLPONDRFLDPEAYGACGARGSPQWVSLNGSLFHCAG 73
DB 2 ARSLPLPQIILLISLALETAGEEAGDKIID---GAPCARSHHWQVALLSGNQLHCGG 57
QY 74 VLVDQSVTLTAHCKNPKLMAVGDHLLLGSEQLRRTTSSVHPKTHQSGPILPRT 133
DB 58 VLAVNRWVLTAAHCKNMEYVHLGSDTLGDRRAQRI-KASKSFRRPGY-----STQT 108
QY 134 DEHDMLKLARPVVPGPRVVALDLPYRCAPGQDQCOYVAGGTAAKVKYKNGITCSSI 193
DB 109 HNDMLVKLNSQALSSWVKVRLPSRCPEPGTTCTVSGMCTTSPDVTPESDLMCVDV 168
QY 194 TILSPKECEVFPVGVVNTNMICAGL-DRGQDPQSDSGPLVCDDETLGILSWGVYPCGS 252
DB 169 KLISPDCTKYKXDLNLSMCAIGIPDSKKNACNGDSGPIVCCRGTLGGLVSWGTFPCGQ 228
QY 253 AQHPAVTQICKYMINKVIRSN 276
DB 229 PNDPGVYTQVCKFTKWINDTWKKH 252

RESULT 13
US-09-210-084-3
Sequence 3, Application US/09210084
Patent No. 6197511
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL KALLIKREIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSES: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09210084-3
```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,084
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/824,874
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0252 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 532504
US-09-210-084-3

Query Match
Best Local Similarity 36.7%; Score 549; DB 3; Length 253;
Matches 108; Conservative 45; Mismatches 97; Indels 14; Gaps 4;

QY 14 ARALAKLPLMAQMAEAALLPONDRFLDPEAYGACGARGSPQWVSLNGSLFHCAG 73
DB 2 ARSLPLPQIILLISLALETAGEEAGDKIID---GAPCARSHHWQVALLSGNQLHCGG 57
QY 74 VLVDQSVTLTAHCKNPKLMAVGDHLLLGSEQLRRTTSSVHPKTHQSGPILPRT 133
DB 58 VLAVNRWVLTAAHCKNMEYVHLGSDTLGDRRAQRI-KASKSFRRPGY-----STQT 108
QY 134 DEHDMLKLARPVVPGPRVVALDLPYRCAPGQDQCOYVAGGTAAKVKYKNGITCSSI 193
DB 109 HNDMLVKLNSQALSSWVKVRLPSRCPEPGTTCTVSGMCTTSPDVTPESDLMCVDV 168
QY 194 TILSPKECEVFPVGVVNTNMICAGL-DRGQDPQSDSGPLVCDDETLGILSWGVYPCGS 252
DB 169 KLISPDCTKYKXDLNLSMCAIGIPDSKKNACNGDSGPIVCCRGTLGGLVSWGTFPCGQ 228
QY 253 AQHPAVTQICKYMINKVIRSN 276
DB 229 PNDPGVYTQVCKFTKWINDTWKKH 252

RESULT 14
US-09-764-762-3
Sequence 3, Application US/09764762
Patent No. 6472195
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL KALLIKREIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSES: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/764,762
```

FILING DATE: 16-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/210,084  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0252 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 532504  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-764-762-3

Query Match 36.7%; Score 549; DB 4; Length 253;  
Best Local Similarity 40.9%; Pred. No. 1,2e-48;  
Matches 108; Conservative 45; Mismatches 97; Indels 14; Gaps 4;

Db 14 ARALAKLPLMLMAQMAEABALLPQNDRLDPEAYGAPCARSGOPWVSLFNGLSFHCAG 73  
2 ARSLLPLQILLSLALSTAGEAEGDKIID---GAPCARSGHPQVALLSGNQLHCG 57  
74 VLDQSVLTAAHGKNGKPIANVGDHLLLOGEQLRTTRSVVHPKXQSGPILPRRT 133  
58 VLVNERVLTAAHCKNNEYVHLSPTLGDRAQR-KASKSFRHPGY-----STGT 108  
134 DEHDLMLKLARPVVPGPRVRAQLPYRCAQPGDQCVAGMGTAAARVYKNGLTCSST 193  
109 HAVDLMLVLKNSQARLSMVKVRLPSRCEPPTCTVSGMGTTSPTDVPFSPDLKCV 168  
194 TILSPCEVFYPGVVNTNMICAGL-DRGQDPQSDSGSPLYCDETLQGLISMGVYPCGS 252  
169 KLISPDCTKRYKDLLENMMLCAGIPDSKKNACNGSGGPLVCRGTLQGLVSWGTFCQ 228  
Qy 194 TILSPCEVFYPGVVNTNMICAGL-DRGQDPQSDSGSPLYCDETLQGLISMGVYPCGS 252  
Db 169 KLISPDCTKRYKDLLENMMLCAGIPDSKKNACNGSGGPLVCRGTLQGLVSWGTFCQ 228  
Qy 253 AQHPAVYTQICKYMSWINKVIRSN 276  
Db 229 PNDGVTQYCKFTKWINDTMKKH 252

RESULT 15  
PCT-US96-04294-2  
Sequence 2, Application PC/TUS9604294  
GENERAL INFORMATION:  
APPLICANT: Dixon, Eric P.  
APPLICANT: Johnstone, Edward M.  
APPLICANT: Little, Sheila P.  
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND  
NUMBER OF SEQUENCES: 3  
TITLE OF INVENTION: RELATED NUCLEIC ACIDS  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/04294

FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/416,257  
FILING DATE: 04-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bialock, Donna K.  
REGISTRATION NUMBER: 38,082  
REFERENCE/DOCKET NUMBER: X9239  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-277-1090  
TELEFAX: 317-276-3861  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 253 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-04294-2

Query Match 36.7%; Score 549; DB 5; Length 253;  
Best Local Similarity 40.9%; Pred. No. 1,2e-48;  
Matches 108; Conservative 45; Mismatches 97; Indels 14; Gaps 4;

Db 14 ARALAKLPLMLMAQMAEABALLPQNDRLDPEAYGAPCARSGOPWVSLFNGLSFHCAG 73  
2 ARSLLPLQILLSLALSTAGEAEGDKIID---GAPCARSGHPQVALLSGNQLHCG 57  
74 VLDQSVLTAAHGKNGKPIANVGDHLLLOGEQLRTTRSVVHPKXQSGPILPRRT 133  
58 VLVNERVLTAAHCKNNEYVHLSPTLGDRAQR-KASKSFRHPGY-----STGT 108  
134 DEHDLMLKLARPVVPGPRVRAQLPYRCAQPGDQCVAGMGTAAARVYKNGLTCSST 193  
109 HAVDLMLVLKNSQARLSMVKVRLPSRCEPPTCTVSGMGTTSPTDVPFSPDLKCV 168  
194 TILSPCEVFYPGVVNTNMICAGL-DRGQDPQSDSGSPLYCDETLQGLISMGVYPCGS 252  
169 KLISPDCTKRYKDLLENMMLCAGIPDSKKNACNGSGGPLVCRGTLQGLVSWGTFCQ 228  
Qy 194 TILSPCEVFYPGVVNTNMICAGL-DRGQDPQSDSGSPLYCDETLQGLISMGVYPCGS 252  
Db 169 KLISPDCTKRYKDLLENMMLCAGIPDSKKNACNGSGGPLVCRGTLQGLVSWGTFCQ 228  
Qy 253 AQHPAVYTQICKYMSWINKVIRSN 276  
Db 229 PNDGVTQYCKFTKWINDTMKKH 252

Search completed: November 25, 2003, 09:09:43  
Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 07:29:59 ; Search time 51 Seconds  
(without alignments)  
858.991 Million cell updates/sec

Title: US-10-021-368-1

Perfect score: 1496  
Sequence: 1 MEAPHLHSAASGAPAAKL.....AVYTCCKMINKVINSN 276

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19jun03:\*

1:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1496	100.0	276	AAW07620	Human NES1 polypep
2	1496	100.0	276	AAAB1327	Human NES1. Homo
3	1496	100.0	276	ABG96356	Human ovarian can
4	1496	100.0	276	AAU82729	Amino acid sequenc
5	729	48.7	133	ABP42175	Human ovarian anti
6	564	37.7	248	AAAB1304	Human Kik-15 prote
7	564	37.7	248	AAAB4032	Human PRO1303 prote
8	564	37.7	248	AAAB4428	Human PRO1303 prote
9	564	37.7	248	AAAY9393	Human PRO1303 (DNO

10	564	37.7	248	AAW23994	Human EST encoded
11	564	37.7	248	AAAB6142	Protein of the inv
12	549	36.7	253	AAAB7888	Human stratum corn
13	549	36.7	253	AAW05383	Human amyloid prec
14	549	36.7	253	ABAB4406	Human SCCE protein
15	549	36.7	253	ABAB4421	Human SCCE protein
16	549	36.7	253	AAU82740	Amino acid sequenc
17	549	36.7	253	ABU07440	Protein different
18	549	36.7	253	ABU07471	Protein different
19	547	36.6	260	AA141744	Human PRO322 prote
20	547	36.6	260	AAAY32852	Human serine prote
21	547	36.6	260	AAAY03220	Amino acid sequenc
22	547	36.6	260	AAAB21322	Human neuropeptin.
23	547	36.6	260	AAAB44300	Human PRO322 (UNO2
24	547	36.6	260	AAAY51131	Human neuropeptin pr
25	547	36.6	260	AAAB23373	Novel human diagno
26	547	36.6	260	AAU12369	Human PRO322 polyp
27	547	36.6	260	AAAB53087	Human angiogenesis
28	547	36.6	260	ABAB94458	Human angiogenesis
29	547	36.6	260	ABAB4852	Human PRO322 prote
30	547	36.6	260	AAU81959	Human PRO322. Hom
31	547	36.6	260	ABU6767	Human PRO polypept
32	547	36.6	260	ABU67043	Human secreted/tira
33	547	36.6	260	ABU59848	Novel secreted and
34	547	36.6	260	ABU61130	Human PRO322 polyp
35	540	35.1	275	AAAB21311	Human neuropeptin.
36	539	36.0	260	AAAY05438	Human protease HUP
37	539	36.0	305	AAAB32853	Human serine prote
38	539	36.0	306	AAAB37985	Human secreted pro
39	539	36.0	315	AAAB41332	Human ovarian anti
40	538	36.0	260	AAAB7703	A human serine pro
41	538	36.0	288	AAAB36482	Fusion gene with h
42	538	36.0	288	AAAB67542	Amino acid sequenc
43	536	35.8	249	ABAB84420	Porcine SCCE prote
44	533	35.6	225	ABAB98502	Human Stratum Corn
45	527	35.2	257	AAAB21326	Human HSCE. Homo

## ALIGNMENTS

RESULT 1	
ID	AAW07620
XX	AAW07620 standard; Protein; 276 AA.
AC	AAW07620;
DT	28-FEB-1997 (first entry)
XX	XX
DE	Human NES1 polypeptide.
XX	XX
KM	NES1; malignancy; cancer; breast carcinoma; cervix carcinoma;
KW	prostate carcinoma; therapy; diagnosis; prognosis; serine protease.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO9639175-A1.
XX	XX
PD	12-DEC-1996.
XX	XX
PF	21-MAY-1996; 96MO-US07343.
XX	XX
PR	06-JUN-1995; 95US-0467455.
XX	XX
PA	(NEMF-) NEW ENGLAND MEDICAL CENT HOSPITALS INC.
XX	XX
PI	Band V;
XX	XX
DR	WPI; 1997-042855/04.
XX	XX
DR	N-FSDS; AAT44111.
XX	XX
PT	NES1 polypeptide, negatively associated with epithelial cell
PT	malignancy - provides diagnostic marker for breast, cervical and

PT prostate carcinoma, and can be useful for creating these diseases  
 XX  
 PS Claim 1; Page 44-45; 77pp; English.

CC NES1 (Normal Epithelial Specific-1) polypeptide (AA007620) is a  
 CC serine protease negatively associated with epithelial cell  
 CC malignancy. The NES1 amino acid sequence was deduced from a cDNA  
 CC clone (A074111) isolated from radiation-transformed human mammary  
 CC epithelial cell line, 76R-30. The polypeptide can be produced in  
 CC transformed host cells. It provides a diagnostic marker for  
 CC breast, cervical and prostate carcinoma and may also be used in  
 CC the treatment of these diseases.

XX Sequence 276 AA;

Query Match 100.0%; Score 1496; DB 18; Length 276;

Best Local Similarity 100.0%; Pred. No. 3.5e-133; Indels 0; Gaps 0;

Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAPHLHLSAASGARALAKLLPLMAQLMAEALPLPNDTRLDPAYGAPCARGSQPMQ 60  
 DB 1 MRAPHLHLSAASGARALAKLLPLMAQLMAEALPLPNDTRLDPAYGAPCARGSQPMQ 60

QY 61 VSLFNGLSFHCAGVLDQSWVLTAAHCNKPMLARVGDHLLLLQGEQLRRTTSVYHPK 120  
 DB 61 VSLFNGLSFHCAGVLDQSWVLTAAHCNKPMLARVGDHLLLLQGEQLRRTTSVYHPK 120

QY 121 YHGGSGPILPRRTDEHDLMLKLARPVVGPVRVALQPYRCAQPGDCCQVAGMGTAAAR 180  
 DB 121 YHGGSGPILPRRTDEHDLMLKLARPVVGPVRVALQPYRCAQPGDCCQVAGMGTAAAR 180

QY 181 RYKYNKGLTCSSTITLSPKECEVFYGVVTNNMICAGLDRGDDPCQSDSGPLVDETLQ 240  
 DB 181 RYKYNKGLTCSSTITLSPKECEVFYGVVTNNMICAGLDRGDDPCQSDSGPLVDETLQ 240

QY 241 GILSWGVPFGCSAQHPAVYTOICKYMSWINKYIRSN 276  
 DB 241 GILSWGVPFGCSAQHPAVYTOICKYMSWINKYIRSN 276

RESULT 2  
 AAB21327

ID AAB21327 standard; Protein; 276 AA.

XX AAB21327;

DT 02-FEB-2001 (first entry)

XX Human NES1.

KM Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; NES1;  
 KM normal epithelial cell-specific 1; kallikrein-like protein;  
 KM serine protease; cytosolic; cancer; prostate cancer.

XX Homo sapiens.

PN WO200053776-A2.

PD 14-SEP-2000.

PF 09-MAR-2000; 2000WO-CA00258.

PR 11-MAR-1999; 99US-0124260.

PR 01-APR-1999; 99US-0127386.

PR 21-UTL-1999; 99US-0144919.

PA (MOUN ) MOUNT SINAI HOSPITAL.

PI Yousef GM, Diamandis EP;

DR WPI; 2000-587440/55.

PT New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L

PT protein mediated disorders, especially cancer.  
 XX  
 PS Example 4; Fig 27; 184pp; English.

CC The present sequence is human normal epithelial cell-specific 1  
 CC (NES1), a member of the serine protease family. Kallikreins and  
 CC kallikrein-like proteins are a subgroup of the serine protease enzyme  
 CC family. They catalyze the selective cleavage of specific polypeptide  
 CC precursors to release peptides with potent biological activity. Nucleic  
 CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,  
 CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the  
 CC treatment, monitoring and diagnosis of cancers, especially prostate  
 CC cancer. They can also be used to identify a substance that can associate  
 CC with or mediate the biological activity of the proteins. Antibodies can  
 CC be used to treat conditions mediated by the kallikrein-like proteins.

XX Sequence 276 AA;

Query Match 100.0%; Score 1496; DB 21; Length 276;

Best Local Similarity 100.0%; Pred. No. 3.5e-133; Indels 0; Gaps 0;

Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAPHLHLSAASGARALAKLLPLMAQLMAEALPLPNDTRLDPAYGAPCARGSQPMQ 60  
 DB 1 MRAPHLHLSAASGARALAKLLPLMAQLMAEALPLPNDTRLDPAYGAPCARGSQPMQ 60

QY 61 VSLFNGLSFHCAGVLDQSWVLTAAHCNKPMLARVGDHLLLLQGEQLRRTTSVYHPK 120  
 DB 61 VSLFNGLSFHCAGVLDQSWVLTAAHCNKPMLARVGDHLLLLQGEQLRRTTSVYHPK 120

QY 121 YHGGSGPILPRRTDEHDLMLKLARPVVGPVRVALQPYRCAQPGDCCQVAGMGTAAAR 180  
 DB 121 YHGGSGPILPRRTDEHDLMLKLARPVVGPVRVALQPYRCAQPGDCCQVAGMGTAAAR 180

QY 181 RYKYNKGLTCSSTITLSPKECEVFYGVVTNNMICAGLDRGDDPCQSDSGPLVDETLQ 240  
 DB 181 RYKYNKGLTCSSTITLSPKECEVFYGVVTNNMICAGLDRGDDPCQSDSGPLVDETLQ 240

QY 241 GILSWGVPFGCSAQHPAVYTOICKYMSWINKYIRSN 276  
 DB 241 GILSWGVPFGCSAQHPAVYTOICKYMSWINKYIRSN 276

RESULT 3  
 ABG96356

ID ABG96356 standard; Protein; 276 AA.

XX ABG96356;

DT 11-DEC-2002 (first entry)

DE Human ovarian cancer marker OV32.

KM Human; ovarian cancer; marker; cancer; familial history; brain disorder;  
 KM central nervous system disorder; bacterial meningitis; viral meningitis;  
 KM Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;  
 KM brain herniation; inflammation; encephalitis; testicular disorder;  
 KM nontuberculous granulomatous orchitis; connective tissue disorder;  
 KM heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;  
 KM histological type; carcinogenic; ovarian cancer marker.

XX Homo sapiens.

PN WO200271928-A2.

PD 19-SEP-2002.

PF 14-MAR-2002; 2002WO-US07826.

PR 14-MAR-2001; 2001US-276025P.

PR 14-MAR-2001; 2001US-276026P.

PR 10-AUG-2001; 2001US-311732P.

PR 19-SEP-2001; 2001US-323580P.

PR 26-SEP-2001; 2001US-324967P.  
 PR 26-SEP-2001; 2001US-325102P.  
 PR 26-SEP-2001; 2001US-325149P.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG,  
 PI Meyer RE, Morrissey MP, Olandt PJ, Sen A, Viohy PO, Mills GB,  
 PI Bast RC, Lu X, Schmandt RE, Zhao X, Glattc K;  
 XX  
 DR WPI; 2002-722277/78.  
 DR N-PSDB; ABS76452.  
 XX  
 PT Assessing whether a patient is afflicted with ovarian cancer, useful in  
 PT assessing the stage or progression of the disease, comprises comparing  
 PT the expression level of a cancer marker in a sample from a patient and  
 PT from a non cancer patient -  
 XX  
 XX  
 PS Disclosure; Page 290; 481pp; English.  
 XX  
 CC The present invention relates to a new method for assessing whether a  
 CC patient is afflicted with ovarian cancer. The method involves comparing  
 CC the expression level of a marker in a patient sample and the normal level  
 CC of expression of the marker in a control non-ovarian cancer sample, where  
 CC the marker is selected from 363 cancer markers described in the  
 CC specification. The method of the invention is useful in diagnosing or  
 CC characterizing cancer, in detecting the presence of cancer as early as  
 CC possible, and the recurrence of ovarian cancer. The method may also be of  
 CC particular use with patients having an enhanced risk of developing  
 CC ovarian cancer (e.g. patients having a familial history of ovarian  
 CC cancer). The cancer markers may be used in the management and treatment  
 CC of e.g. brain and central nervous system disorders (e.g. bacterial and  
 CC viral meningitis, Alzheimer's disease or Parkinson's disease), brain  
 CC disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),  
 CC inflammations (e.g. bacterial or viral meningitis or encephalitis),  
 CC testicular disorders (e.g. nontuberculous granulomatous orchitis),  
 CC connective tissue disorders, or heart disorders (e.g. ischaemic heart  
 CC disease or atherosclerosis). The compositions and methods may also be  
 CC used in assessing the histological type of neoplasm associated with  
 CC ovarian cancer, monitoring the progression of ovarian cancer,  
 CC determining whether ovarian cancer has metastasized or is likely to  
 CC metastasize, selecting a composition for inhibiting ovarian cancer,  
 CC assessing the ovarian carcinogenic potential of a compound, or  
 CC inhibiting ovarian cancer or at risk of developing ovarian cancer. The  
 CC present amino acid sequence represents one of the ovarian cancer markers  
 CC described in the invention.  
 CC  
 XX  
 XX  
 SQ Sequence 276 AA;  
 Query Match 100.0%; Score 1496; DB 23; Length 276;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-133;  
 Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRAPIHLISAAGARALAKLPLMLAQLMAABALLPQNDTRLDPEAYGAPCARGSQPMQ 60  
 DB 1 MRAPIHLISAAGARALAKLPLMLAQLMAABALLPQNDTRLDPEAYGAPCARGSQPMQ 60  
 QY 61 VSLFNGLSHFACAGLVLDQSWVLTAAHCGNKPIMARVGDHLLLLQGEQLRTTRSVYHAK 120  
 DB 61 VSLFNGLSHFACAGLVLDQSWVLTAAHCGNKPIMARVGDHLLLLQGEQLRTTRSVYHAK 120  
 QY 121 YHOGSGPILFRRTDEHDMILKLARPVPGPRVALQLPFRCAQPGDCCVAGMGTTAAR 180  
 DB 121 YHOGSGPILFRRTDEHDMILKLARPVPGPRVALQLPFRCAQPGDCCVAGMGTTAAR 180  
 QY 181 RYKXNKGLTSSITLISPEKCEVFYPGVNTNMICAGLDGQPPCCSDSGGPLVCEBTLQ 240  
 DB 181 RYKXNKGLTSSITLISPEKCEVFYPGVNTNMICAGLDGQPPCCSDSGGPLVCEBTLQ 240  
 QY 241 GILSMGVYPCGSAQHAPAVYTOICKYMSWINKVIRSN 276  
 DB 241 GILSMGVYPCGSAQHAPAVYTOICKYMSWINKVIRSN 276

RESULT 4  
 ID AAU82729 standard; Protein; 276 AA.  
 XX  
 AC AAU82729;  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Amino acid sequence of novel human protease #28.  
 XX  
 KW Human; protease; cancer; immune-related disorder; cardiovascular disease;  
 KW neuronal-associated disease; metabolic disorder; inflammatory disorder;  
 KW nervous system disorder; sexual dysfunction; pain; mood disorder;  
 KW hypertension; psychotic disorder; neurological disorder; dyskinesia;  
 KW viral infection; human immunodeficiency virus; HIV; non-viral infection;  
 KW ocular disease; cytostatic; enzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200200860-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 26-JUN-2001; 2001WO-US20171.  
 XX  
 PR 26-JUN-2000; 2000US-214047P.  
 XX  
 PA (SUGEN-) SUGEN INC.  
 XX  
 PI Plozman G, Whyte D, Sudarsanam S, Manning G, Caenepeel S;  
 PI Charyczak G;  
 XX  
 DR WPI; 2002-139913/18.  
 DR N-PSDB; ABK31771.  
 XX  
 PT Nucleic acids encoding novel human proteases, useful for useful for  
 PT treating diseases and disorders such as cancers, immune-related  
 PT diseases and disorders, cardiovascular disease (e.g. stenosis) and  
 PT inflammatory disorders -  
 XX  
 PS Claim 6; Fig 2K; 313pp; English.  
 XX  
 CC The present invention relates to the isolation of novel human  
 CC proteases, and the nucleic acids encoding them. The sequences of  
 CC the invention are useful for treating diseases and disorders such as  
 CC cancers (e.g. breast, colon, lung), immune-related diseases and disorders  
 CC (e.g. inflammatory diseases and asthma), cardiovascular diseases  
 CC (e.g. restenosis and coronary thrombosis), brain or neuronal-associated  
 CC diseases, metabolic disorders (e.g. diabetes, obesity), inflammatory  
 CC disorders (e.g. rheumatoid arthritis and psoriasis), central or  
 CC peripheral nervous system diseases, migraines, pain, sexual dysfunction,  
 CC mood disorders, attention disorders, cognition disorders, hypertension,  
 CC hypertension, psychotic disorders, neurological disorders  
 CC (e.g. Alzheimer's disease, Parkinson's disease) and dyskinesias.  
 CC The nucleic acids and polypeptides are also useful for treating viral  
 CC infections caused by human immunodeficiency virus (HIV), and non-viral  
 CC infections such as ocular disease (e.g. glaucoma) and macular  
 CC degeneration. AAU82702-AAU82760 represent the novel human proteases of  
 CC the invention.  
 CC  
 XX  
 XX  
 SQ Sequence 276 AA;  
 Query Match 100.0%; Score 1496; DB 23; Length 276;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-133;  
 Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRAPIHLISAAGARALAKLPLMLAQLMAABALLPQNDTRLDPEAYGAPCARGSQPMQ 60  
 DB 1 MRAPIHLISAAGARALAKLPLMLAQLMAABALLPQNDTRLDPEAYGAPCARGSQPMQ 60  
 QY 61 VSLFNGLSHFACAGLVLDQSWVLTAAHCGNKPIMARVGDHLLLLQGEQLRTTRSVYHAK 120  
 DB 61 VSLFNGLSHFACAGLVLDQSWVLTAAHCGNKPIMARVGDHLLLLQGEQLRTTRSVYHAK 120

Db 61 VSLFNGISFHCAGVLDVDSWVLPRAHGKFLMARVGDHLLLLQSEQLRRTTRSVVHPK 120  
 QY 121 YHQSSEGLPRRDEHDLMLKLARPVPGPRVPAIQLPYRCAPQPDQCVAGMGTAAAR 180  
 Db 121 YHQSSEGLPRRDEHDLMLKLARPVPGPRVPAIQLPYRCAPQPDQCVAGMGTAAAR 180  
 QY 181 RVKYNKGLTSSGTTILSPKCEVFYPCVVTNNMTICGLDGGDPCQSDSGSPVCDITLQ 240  
 Db 181 RVKYNKGLTSSGTTILSPKCEVFYPCVVTNNMTICGLDGGDPCQSDSGSPVCDITLQ 240  
 QY 241 GILSMGVYPCGSAQHPAVYTOICKYSWINKVIRSN 276  
 Db 241 GILSMGVYPCGSAQHPAVYTOICKYSWINKVIRSN 276  
 RESULT 5  
 ID ABP42175 standard; Protein; 133 AA.  
 AC ABP42175;  
 DT 22-AUG-2002 (first entry)  
 DE Human ovarian antigen HCOPE27, SEQ ID NO:3307.  
 XX Human, ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KM Human cancer; breast cancer; tumour; reproductive system disorder;  
 KM Infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KM PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KM Inflammatory condition; immune disorder; blood disorder;  
 KM Cardiovascular disorder; respiratory disorder; neurological disorder;  
 KM Gastrointestinal disorder; urinary system disorder; drug screening;  
 KM Gene therapy; chromosome mapping; forensic analysis;  
 KM Antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KM Antiinflammatory; gynaecological; reproductive.  
 OS Homo sapiens.  
 XX WO200200677-A1.  
 PN 03-JAN-2002.  
 PD 07-JUN-2001; 2001WO-US18569.  
 PF 07-JUN-2000; 2000US-209467P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Birst CE, Rosen CA;  
 PI Birst CE, Rosen CA;  
 XX WPI; 2002-147878/19.  
 DR N-PSDB; ABQ55252.  
 XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
 PT ovarian cancer), immune disorders, cardiovascular disorders and  
 PT neurological diseases -  
 PS Claim 11; SEQ ID NO 3307; 2922pp; English.  
 XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovary and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic

CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents a human ovarian antigen of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 133 AA;  
 XX Query Match 48.7%; Score 729; DB 23; Length 133;  
 XX Best local Similarity 99.2%; Pred. No. 4,6e-61;  
 XX Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 144 ARPVPFGRVRAIQLPYRCAPQPDQCVAGMGTAAARVKNKGLTSSGTTILSPKCEV 203  
 Db 1 ARPVPFGRVRAIQLPYRCAPQPDQCVAGMGTAAARVKNKGLTSSGTTILSPKCEV 60  
 QY 204 FYPGVVTNNMTICGLDGGDPCQSDSGSPVCDITLQGISWGVYPCGSAQHPAVYTOIC 263  
 Db 61 FYPGVVTNNMTICGLDGGDPCQSDSGSPVCDITLQGISWGVYPCGSAQHPAVYTOIC 120  
 QY 264 KYMSWINKVIRSN 276  
 Db 121 KYMSWINKVIRSN 133  
 RESULT 6  
 ID AAB21304 standard; Protein; 248 AA.  
 AC AAB21304;  
 DT 02-FEB-2001 (first entry)  
 DE Human KLR-L5 protein #4.  
 XX Human; KLR-L1; KLR-L2; KLR-L3; KLR-L4; KLR-L5; KLR-L6;  
 KM Kallikrein-like protein; serine protease; cytostatic; cancer;  
 KM prostate cancer.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX WO200053776-A2.  
 PN 14-SEP-2000.  
 PD 09-MAR-2000; 2000WO-CA00258.  
 PF 11-MAR-1999; 99US-0124260.  
 PR 01-APR-1999; 99US-0127386.  
 PR 21-JUL-1999; 99US-0144919.  
 XX (MOUN ) MOUNT SINAI HOSPITAL.  
 PA Yousef GM, Diamandis EP;  
 PI Yousef GM, Diamandis EP;  
 XX WPI; 2000-587440/55.  
 DR N-PSDB; AAA95944.  
 XX New kallikrein-like (KLR-L) proteins for diagnosing and treating KLR-L  
 PT protein mediated disorders, especially cancer.  
 PS Claim 12; Page 172; 184pp; English.



XX The present sequence is one of four alternatively spliced kallikrein-like  
 CC proteins encoded by the human KLK-L4 gene. Kallikreins and  
 CC kallikrein-like proteins are a subgroup of the serine protease enzyme  
 CC family. They catalyse the selective cleavage of specific polypeptide  
 CC precursors to release peptides with potent biological activity. Nucleic  
 CC acid encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,  
 CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the  
 CC treatment, monitoring and diagnosis of cancers, especially prostate  
 CC cancer. They can also be used to identify a substance that can associate  
 CC with or mediate the biological activity of the proteins. Antibodies can  
 CC be used to treat conditions mediated by the kallikrein-like proteins.

XX SQ Sequence 248 AA;

Query Match 37.7%; Score 564; DB 21; Length 248;  
 Best Local Similarity 45.9%; Pred. No. 3.9e-45;  
 Matches 118; Conservative 29; Mismatches 92; Indels 18; Gaps 5;

QY 23 LLMAGLMAEAPLLPQNDTRLDPEAY-GAPCARSGSQPMQVSLFNGLSFHCAGVLVDQSNV 81  
 DB 7 LLLCVLGLSQAA-----TPKIFNGTEGGRNSQPMQVGLFEGTSLRCGGVLDHKNV 57  
 QY 82 LTTAAHCGNKPMLARVGDHLLLIQ-GEQLRRTTRSVVHPKTHQSGGPILPRTDEHIDL 140  
 DB 58 LTTAAHCGSRVYVRLGSHLSQLDWTETQIRHSGFSVTHPGVLGAS-----TSHEDHIDL 111  
 QY 141 LKTLARPVVPRVRLALQLPYRCAPGDPCCQVAGMGTTAARVKNYKGLTCSITLSPKE 200  
 DB 112 LRLRLPVRVTVSSVQPLPLPNDCAATGTECHVSGMGITTHNPFPDLQLCLNLISIVSHAT 171  
 QY 201 CEVFPYGVVTVNNMTCAGLDRGDPCQSDSGGFLVCDETLGILSWG-VYPCGSAQHPRVY 259  
 DB 172 CHGVYPRGRITSNMVCAGVPGQDACQSDSGGFLVCGVLQGLVSGVSGVPGQDGIPEVY 231  
 QY 260 TQICKYMSWINKVIRSN 276  
 DB 232 TYICKYVDWIRMIRNN 248

RESULT 7  
 AAB24032  
 ID AAB24032 standard; Protein; 248 AA.

XX AAB24032;  
 XX 25-JAN-2001 (first entry)  
 XX  
 DE Human PRO1303 protein sequence SEQ ID NO:33.  
 XX  
 KW Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;  
 KW identification; tumorigenesis; anticancer; detection.

XX OS Homo sapiens.  
 XX PN MO200053750-A1.  
 XX 14-SEP-2000.

XX PD 14-SEP-2000.  
 XX PF 02-DEC-1999; 99WO-US28551.  
 XX PR 08-MAR-1999; 99WO-US05028.  
 XX PR 01-SEP-1999; 99WO-US20111.  
 XX PR 29-OCT-1999; 99US-0162506.  
 XX PR 30-NOV-1999; 99WO-US28313.  
 XX PR 01-DEC-1999; 99WO-US28634.

XX PA (GETH ) GENENTECH INC.

XX PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;  
 XX WPI; 2000-594320/56.  
 XX DR N-PSDB; AAC58114.

XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit  
 PT the growth of tumors in mammals, and to identify inhibitors of PRO  
 PT polypeptide activity or expression.  
 XX  
 PS Claim 61; Fig 24; 226pp; English.

XX The present invention describes an antibody that binds to a human  
 CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;  
 CC PRO3439; PRO1927; PRO3567; PRO1295; PRO1093; PRO4344; PRO4354;  
 CC PRO4397; PRO4007; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has  
 CC anticancer activity and can be used to diagnose tumours in mammals, by  
 CC detecting complex formation when the antibody is contacted with test  
 CC cells. Increased expression of genes encoding (I) can also be detected  
 CC to diagnose tumours. Agents which inhibit the activity of (I),  
 CC especially the antibodies, or an antisense oligonucleotide which  
 CC hybridises to genes encoding (I), can be used to inhibit tumour growth,  
 CC preferably by inducing cell death. Methods from the present invention  
 CC can be used to identify compounds which inhibit the biological activity  
 CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation  
 CC probes used in examples from the present invention for human PRO  
 CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human  
 CC PRO polynucleotide and protein sequences given in the exemplification of  
 CC the present invention.

XX SQ Sequence 248 AA;

Query Match 37.7%; Score 564; DB 21; Length 248;  
 Best Local Similarity 45.9%; Pred. No. 3.9e-45;  
 Matches 118; Conservative 29; Mismatches 92; Indels 18; Gaps 5;

QY 23 LLMAGLMAEAPLLPQNDTRLDPEAY-GAPCARSGSQPMQVSLFNGLSFHCAGVLVDQSNV 81  
 DB 7 LLLCVLGLSQAA-----TPKIFNGTEGGRNSQPMQVGLFEGTSLRCGGVLDHKNV 57  
 QY 82 LTTAAHCGNKPMLARVGDHLLLIQ-GEQLRRTTRSVVHPKTHQSGGPILPRTDEHIDL 140  
 DB 58 LTTAAHCGSRVYVRLGSHLSQLDWTETQIRHSGFSVTHPGVLGAS-----TSHEDHIDL 111  
 QY 141 LKTLARPVVPRVRLALQLPYRCAPGDPCCQVAGMGTTAARVKNYKGLTCSITLSPKE 200  
 DB 112 LRLRLPVRVTVSSVQPLPLPNDCAATGTECHVSGMGITTHNPFPDLQLCLNLISIVSHAT 171  
 QY 201 CEVFPYGVVTVNNMTCAGLDRGDPCQSDSGGFLVCDETLGILSWG-VYPCGSAQHPRVY 259  
 DB 172 CHGVYPRGRITSNMVCAGVPGQDACQSDSGGFLVCGVLQGLVSGVSGVPGQDGIPEVY 231  
 QY 260 TQICKYMSWINKVIRSN 276  
 DB 232 TYICKYVDWIRMIRNN 248

RESULT 8  
 AAB24428  
 ID AAB24428 standard; Protein; 248 AA.

XX AAB24428;  
 XX 07-NOV-2000 (first entry)  
 XX  
 DE Human PRO1303 protein sequence SEQ ID NO:203.

XX KW Human; PRO; promotion; inhibition; angiogenesis; cardiovascularisation;  
 KW diagnosis; trauma; wound; cancer; atherosclerosis; cardiac hypertrophy;  
 KW angiogenic; proliferative; cardiac; cardiovascular; antiatherosclerotic;  
 KW cytosclastic; gene therapy; vaccine.

XX OS Homo sapiens.

XX PN MO200032221-A2.  
 XX 08-JUN-2000.

PF 30-NOV-1999; 99WO-US28311.  
 XX 01-DEC-1998; 98WO-US25108.  
 PR 15-DEC-1998; 98US-012850.  
 PR 12-JAN-1999; 99US-011554.  
 PR 08-MAR-1999; 99WO-US05028.  
 PR 12-MAR-1999; 99US-012395.  
 PR 28-APR-1999; 99US-013145.  
 PR 14-MAY-1999; 99US-013428.  
 PR 02-JUN-1999; 99WO-US12252.  
 PR 23-JUN-1999; 99US-014103.  
 PR 20-JUL-1999; 99US-014475.  
 PR 26-YUL-1998; 99US-014568.  
 PR 01-SEP-1998; 99WO-US20111.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 13-SEP-1999; 99WO-US20944.  
 PR 15-SEP-1999; 99WO-US21090.  
 PR 15-SEP-1999; 99WO-US21547.  
 PR 05-OCT-1999; 99WO-US21089.  
 PR 29-OCT-1999; 99US-0162506.  
 XX (GETH) GENENTECH INC.  
 XX Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Hillan KJ, Goddard A;  
 PI Godowski PJ, Gurney AL, Klein RD, Kuo SS, Paoni NF, Smith V;  
 PI Watanabe CK, Williams PM, Wood WI;  
 XX N-PSDB; AAA77671.  
 DR MPI: 2000-412154/35.  
 XX

XX Nucleic acids encoding PRO polypeptides useful for preventing,  
 PT diagnosing and treating disorders a cardiovascular, endothelial or  
 PT angiogenic disorders in mammals -  
 XX  
 PS Claim 72; Fig 82; 315BP; English.  
 XX

XX The present invention describes nucleic acids encoding PRO polypeptides  
 CC useful for preventing, diagnosing and treating disorders a  
 CC cardiovascular, endothelial or angiogenic disorder in mammals by  
 CC modulating cell proliferation, angiogenesis and cardiovascularisation,  
 CC and for identifying agonists and antagonists of these processes. The  
 CC nucleic acids and the proteins they encode may be used in the  
 CC prevention, treatment and diagnosis of diseases associated with  
 CC inappropriate PRO expression such as cardiovascular, endothelial or  
 CC angiogenic disorders in mammals (e.g. atherosclerosis, cancers and  
 CC cardiac hypertrophy). For example, the nucleic acids (NAs) and vectors  
 CC containing them and the PRO polypeptide may be used to treat disorders  
 CC associated with decreased PRO expression. AAA77510 to AAA77721 and  
 CC AAB24388 to AAB24435 represent nucleotide and protein sequences used in  
 CC the exemplification of the present invention.  
 XX

XX Sequence 248 AA;

Query Match 37.7%; Score 564; DB 21; Length 248;  
 Best Local Similarity 45.9%; Fred. No. 3.9e-45;  
 Matches 118; Conservative 29; Mismatches 92; Indels 18; Gaps 5;

QY 23 LLMAGLMAAEALALQNDTRLDPKAY-GAPCARSGOPMVSLENGEFGHCGVVDGWSV 81  
 DB 7 LLLCVGLGSLQRA-----TPKIFNTECGKNSQPMQVLEFGTILRCGGLIDHRWV 57  
 QY 82 LTFNACGKPKLMAVYGDHLLLLQ-GEQLARTTSVNAFKYHQSGETLPRTDDEHML 140  
 DB 58 LTFNACGSGRYVVRLEHSLSQLDWTQIRHSGFSVTPGLGAS-----TSHEDURL 111  
 QY 141 LKLAAPVVPGRVRLQIPYCAQPCQCOVAGNCTTAPARVKYKNKLTCGSTITLSPE 200  
 DB 112 LKLRPVAVTSSVQPLPNDCAITGTCCHSGMGTINHPNPPDLLQCLNLSIVSHAT 171  
 QY 201 CEVFPYGVVNTNMICAGLDRGQPCQSDSGSPVYCDFTLLQSLISWG-VYPCGSAQHPAVY 259  
 DB 172 CHGVVFGRTSNMVCAGGVFGQDAGCGDSGGPLVCGGVLLQCLVMSGVSVPCCQDGPVGY 231

QY 260 TQICKYMSWINKVIRSN 276  
 DB 232 TYICKYVDWIRMIMENN 248

# RESULT 9

AAV99393  
 ID AAV99393 standard; Protein; 248 AA.

AAV99393;

08-ANG-2000 (first entry)

Human PRO1303 (UNC669) amino acid sequence SEQ ID NO:194.

Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;  
 transmembrane; secretion; immunoadhesion; pharmaceutical; screening.

Homo sapiens.

MO200012708-A2.

09-MAR-2000.

01-SEP-1999; 99WO-US20111.

01-SEP-1998; 98US-0098716.

01-SEP-1998; 98US-0098749.

02-SEP-1998; 98US-0098750.

02-SEP-1998; 98US-0098803.

02-SEP-1998; 98US-0098821.

09-SEP-1998; 98US-0098843.

09-SEP-1998; 98US-0098956.

09-SEP-1998; 98US-0098958.

09-SEP-1998; 98US-0098962.

10-SEP-1998; 98US-00989741.

10-SEP-1998; 98US-00989754.

10-SEP-1998; 98US-00989763.

10-SEP-1998; 98US-00989808.

10-SEP-1998; 98US-00989812.

10-SEP-1998; 98US-00989815.

15-SEP-1998; 98US-00989816.

15-SEP-1998; 98US-0100385.

15-SEP-1998; 98US-0100388.

16-SEP-1998; 98US-0100390.

16-SEP-1998; 98US-0100584.

16-SEP-1998; 98US-0100627.

16-SEP-1998; 98US-0100631.

16-SEP-1998; 98US-0100662.

16-SEP-1998; 98US-0100664.

17-SEP-1998; 98US-0100683.

17-SEP-1998; 98US-0100684.

17-SEP-1998; 98US-0100710.

17-SEP-1998; 98US-0100711.

17-SEP-1998; 98US-0100919.

17-SEP-1998; 98US-0100930.

18-SEP-1998; 98US-0100848.

18-SEP-1998; 98US-0100849.

18-SEP-1998; 98US-0101014.

18-SEP-1998; 98US-0101068.

18-SEP-1998; 98US-0101071.

22-SEP-1998; 98US-0101279.

23-SEP-1998; 98US-0101471.

23-SEP-1998; 98US-0101472.

23-SEP-1998; 98US-0101474.

23-SEP-1998; 98US-0101475.

23-SEP-1998; 98US-0101476.

23-SEP-1998; 98US-0101477.

23-SEP-1998; 98US-0101479.

24-SEP-1998; 98US-0101738.

PR 24-SEP-1998; 98US-0101741.  
 PR 24-SEP-1998; 98US-0101743.  
 PR 24-SEP-1998; 98US-0101915.  
 PR 24-SEP-1998; 98US-0101916.  
 PR 29-SEP-1998; 98US-0102207.  
 PR 29-SEP-1998; 98US-0102240.  
 PR 29-SEP-1998; 98US-0102307.  
 PR 29-SEP-1998; 98US-0102330.  
 PR 30-SEP-1998; 98US-0102331.  
 PR 30-SEP-1998; 98US-0102484.  
 PR 30-SEP-1998; 98US-0102487.  
 PR 30-SEP-1998; 98US-0102570.  
 PR 01-OCT-1998; 98US-0102571.  
 PR 01-OCT-1998; 98US-0102684.  
 PR 02-OCT-1998; 98US-0102687.  
 PR 02-OCT-1998; 98US-0102965.  
 PR 06-OCT-1998; 98US-0103258.  
 PR 06-OCT-1998; 98US-0103449.  
 PR 07-OCT-1998; 98US-0103314.  
 PR 07-OCT-1998; 98US-0103315.  
 PR 07-OCT-1998; 98US-0103328.  
 PR 07-OCT-1998; 98US-0103395.  
 PR 07-OCT-1998; 98US-0103396.  
 PR 07-OCT-1998; 98US-0103401.  
 PR 08-OCT-1998; 98US-0103633.  
 PR 08-OCT-1998; 98US-0103678.  
 PR 08-OCT-1998; 98US-0103679.  
 PR 08-OCT-1998; 98US-0103711.  
 PR 14-OCT-1998; 98US-0104257.  
 PR 20-OCT-1998; 98US-0104987.  
 PR 20-OCT-1998; 98US-0105000.  
 PR 20-OCT-1998; 98US-0105002.  
 PR 21-OCT-1998; 98US-0105104.  
 PR 22-OCT-1998; 98US-0105169.  
 PR 22-OCT-1998; 98US-0105266.  
 PR 26-OCT-1998; 98US-0105693.  
 PR 26-OCT-1998; 98US-0105694.  
 PR 27-OCT-1998; 98US-0105807.  
 PR 27-OCT-1998; 98US-0105881.  
 PR 27-OCT-1998; 98US-0105882.  
 PR 27-OCT-1998; 98US-0106062.  
 PR 28-OCT-1998; 98US-0106062.  
 PR 28-OCT-1998; 98US-0106023.  
 PR 28-OCT-1998; 98US-0106029.  
 PR 28-OCT-1998; 98US-0106030.  
 PR 28-OCT-1998; 98US-0106032.  
 PR 28-OCT-1998; 98US-0106033.  
 PR 28-OCT-1998; 98US-0106178.  
 PR 29-OCT-1998; 98US-0106248.  
 PR 29-OCT-1998; 98US-0106384.  
 PR 29-OCT-1998; 98US-0106500.  
 PR 30-OCT-1998; 98US-0106464.  
 PR 03-NOV-1998; 98US-0106856.  
 PR 03-NOV-1998; 98US-0106902.  
 PR 03-NOV-1998; 98US-0106905.  
 PR 03-NOV-1998; 98US-0106919.  
 PR 03-NOV-1998; 98US-0106932.  
 PR 03-NOV-1998; 98US-0106934.  
 PR 10-NOV-1998; 98US-0107783.  
 PR 17-NOV-1998; 98US-0108775.  
 PR 17-NOV-1998; 98US-0108779.  
 PR 17-NOV-1998; 98US-0108787.  
 PR 17-NOV-1998; 98US-0108788.  
 PR 17-NOV-1998; 98US-0108801.  
 PR 17-NOV-1998; 98US-0108802.  
 PR 17-NOV-1998; 98US-0108806.  
 PR 17-NOV-1998; 98US-0108807.  
 PR 17-NOV-1998; 98US-0108867.  
 PR 17-NOV-1998; 98US-0108925.  
 PR 18-NOV-1998; 98US-0108848.  
 PR 18-NOV-1998; 98US-0108849.  
 PR 18-NOV-1998; 98US-0108850.  
 PR 18-NOV-1998; 98US-0108851.  
 PR 18-NOV-1998; 98US-0108852.

PR 18-NOV-1998; 98US-0108858.  
 PR 18-NOV-1998; 98US-0108904.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;  
 XX WPI, 2000-237871/20.  
 DR N-PSDB; AAA37075.  
 XX  
 PT New mammalian DNA sequences encoding transmembrane, receptor or  
 PT secreted PRO polypeptides, useful for screening of potential peptide or  
 PT small molecule inhibitors of the relevant receptor/ligand interactions  
 XX  
 PS Claim 12, Fig 106; 773pp; English.  
 XX  
 CC AAA37022 to AAA37144 encode the new isolated human transmembrane,  
 CC receptor or secreted PRO polypeptides given in AA99340 to AA99463. The  
 CC transmembrane and receptor PRO proteins can be used for screening of  
 CC potential peptide or small molecule inhibitors of the relevant  
 CC receptor/ligand interactions. The polypeptides and nucleotide sequences  
 CC encoding them have various industrial applications, including uses as  
 CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent  
 CC PCR primers and hybridisation probes used in the isolation of the PRO  
 CC polypeptides from the present invention.  
 XX  
 SQ Sequence 248 AA;  
 Query Match 37.7%; Score 564; DB 21; Length 248;  
 Best Local Similarity 45.9%; Pred. No. 3,9e-45;  
 Matches 118; Conservative 29; Mismatches 92; Indels 18; Gaps 5;  
 QY 23 LHMQLMAAEALLPONDRLDEPAY-GAPCARSGSQPMQVSLFNGLSFHCAGVLVDQSW 81  
 Db 7 LILCVLGLGSA-----TPKIFNGTECGRNSQPMQVSLFNGLSFHCAGVLVDHHRV 57  
 QY 82 LTAACGKPLMARVGDHLLIQ-GEOLARTTSVYHPKHKQSGPILPRRTEHDML 140  
 Db 58 LTAACGSRKYVAIGHSLSQDMTEIRISGSSVTHFGTAS-----TSHEDRL 111  
 QY 141 LKLAPVPGPRVALQLPYRCAQPGDCCQYAGWGTAAAREVKYKGLTCSITILLSPKE 200  
 Db 112 LRLRLPVRTSSVQPLPLPNDCAATAGTECHVSGMGITNHPNPEDLLQCLNLSIVSHAT 171  
 QY 201 CEVFPYGVYNNMTCAIGDRGDPDQSDSGPFLVCDLTLQILMNG-VYFPGSAHPVY 259  
 Db 172 CHGVYPRITSNMWCAGGVPGDACCQDSGGPLVCGGVLCGLVSMGSGVPGGDGIPGVY 231  
 QY 260 TQICKYMSWINKVIRSN 276  
 Db 232 TYICKYVDWIMIRNN 248  
 RESULT 10  
 AAM23994  
 ID AAM23994 standard; Protein; 248 AA.  
 XX  
 AC AAM23994;  
 XX  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Human EST encoded protein SEQ ID NO: 1519.  
 XX  
 KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
 KW diagnostics; forensic test; gene mapping; genetic disorder;  
 KW biodiversity; gene therapy; nutrition.  
 XX  
 OS Homo sapiens.  
 XX  
 EN MO200154477-A2.  
 XX  
 PD 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02687.  
 XX  
 XX 25-JAN-2000; 2000US-0491404.  
 PR 17-JUL-2000; 2000US-0617746.  
 PR 03-AUG-2000; 2000US-0631451.  
 PR 15-SEP-2000; 2000US-0663870.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
 PI Cao Y, Drmanac RA, Zhang J, Wehrman T;  
 DR WPI; 2001-476164/51.  
 DR N-PSDB; AAH98653.  
 XX  
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 XX  
 XX Claim 20; Page 1048-1049; 1275pp; English.  
 XX  
 XX The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a  
 CC protein of the invention.  
 XX  
 XX  
 SQ Sequence 248 AA;  
 Query Match 37.7%; Score 564; DB 22; Length 248;  
 Best Local Similarity 45.9%; Pred. No. 3.9e-45;  
 Matches 118; Conservative 29; Mismatches 92; Indels 18; Gaps 5;  
 QY 23 LMAQMAAEALALPQNDTRLDEAY-GAPCARGSQPMQVSLFNGSLFHCAVLVDQSV 81  
 DB 7 LILCVLGLSQA-----TPKIFNGTECGNSQPMQVGLFEGTSLRCGVLIDHRV 57  
 QY 82 LTAAGCNKPLMARVGDHLLLLQ-GEOLRRTTSVAPKXHQSGPILPRRTDEHDLML 140  
 DB 58 LTAAGCGSRVWRLGHEHSLQLDWTETQIRHSGFSVTHRGVLAGS-----TSHEDDL 111  
 QY 141 LKLARVVPGRPRALQLPYRCAGQPDQCOVAGMGTARAVKYNKGLTCSITLSPKE 200  
 DB 112 LRLRLPVRTSSVQPLPNDCAITAGTECHSGMGTINHRNRPDPLQCLNLISVSHAT 171  
 QY 201 CEVYPGVVTNNMTCAGLDRGQDPCQSDSGPLVCDDELQILSMG-VYPCGSAQHNAVY 259  
 DB 172 CHGVYFGRITSNMVCAGGVPGQACQDGGGGLVCGVLAGLVMSGSGVPCQDGIPIGVY 231  
 QY 260 TQICKTMSMINKVIRSN 276  
 DB 232 TYICKYVDWIMIRNN 248

RESULT 11  
 AAB66142  
 ID AAB66142 standard; protein; 248 AA.  
 XX  
 XX AAB66142;  
 AC  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Protein of the invention #54.  
 XX  
 XX Secreted; transmembrane; gene therapy.  
 OS  
 XX Unidentified.  
 XX  
 XX MO2000078961-A1.

PD 28-DEC-2000.  
 XX  
 XX 18-FEB-2000; 2000WO-US04342.  
 XX  
 XX 23-JUN-1999; 99US-0141037.  
 PR 20-JUL-1999; 99US-0144758.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00376.  
 XX  
 XX (GETH) GENENTECH INC.  
 XX  
 XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;  
 PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
 PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;  
 PI Matanabe CK, Williams PM, Wood WI;  
 DR WPI; 2001-071395/08.  
 XX  
 XX Secreted and transmembrane proteins and nucleic acids designated PRO,  
 PT useful as hybridization probes, in chromosome and gene mapping and gene  
 PT therapy -  
 XX  
 XX Claim 1; Fig 108; 787pp; English.  
 XX  
 XX The present invention relates to secreted and transmembrane proteins.  
 CC These proteins and the DNA encoding them may be used as hybridization  
 CC probes, in chromosome and gene mapping and in the generation of  
 CC anti-sense RNA and DNA. They may also be used to generate either  
 CC transgenic animals or knockout animals which are in turn useful for  
 CC development and screening of therapeutically useful reagents.  
 CC The nucleic acids may also be used in gene therapy.  
 XX  
 XX  
 SQ Sequence 248 AA;  
 Query Match 37.7%; Score 564; DB 22; Length 248;  
 Best Local Similarity 45.9%; Pred. No. 3.9e-45;  
 Matches 118; Conservative 29; Mismatches 92; Indels 18; Gaps 5;  
 QY 23 LMAQMAAEALALPQNDTRLDEAY-GAPCARGSQPMQVSLFNGSLFHCAVLVDQSV 81  
 DB 7 LILCVLGLSQA-----TPKIFNGTECGNSQPMQVGLFEGTSLRCGVLIDHRV 57  
 QY 82 LTAAGCNKPLMARVGDHLLLLQ-GEOLRRTTSVAPKXHQSGPILPRRTDEHDLML 140  
 DB 58 LTAAGCGSRVWRLGHEHSLQLDWTETQIRHSGFSVTHRGVLAGS-----TSHEDDL 111  
 QY 141 LKLARVVPGRPRALQLPYRCAGQPDQCOVAGMGTARAVKYNKGLTCSITLSPKE 200  
 DB 112 LRLRLPVRTSSVQPLPNDCAITAGTECHSGMGTINHRNRPDPLQCLNLISVSHAT 171  
 QY 201 CEVYPGVVTNNMTCAGLDRGQDPCQSDSGPLVCDDELQILSMG-VYPCGSAQHNAVY 259  
 DB 172 CHGVYFGRITSNMVCAGGVPGQACQDGGGGLVCGVLAGLVMSGSGVPCQDGIPIGVY 231  
 QY 260 TQICKTMSMINKVIRSN 276  
 DB 232 TYICKYVDWIMIRNN 248

RESULT 12  
 AAR67888  
 ID AAR67888 standard; Protein; 253 AA.  
 XX  
 XX AAR67888;  
 AC  
 XX  
 DT 25-MAR-2003 (updated)  
 XX  
 XX 09-AUG-1995 (first entry)

XX DE Human stratum corneum chymotryptic recombinant enzyme (SCCE).  
XX KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;  
XX KM callosities; keratosis pilaris; ichthyoses; eczema.  
XX OS Homo sapiens.  
XX PN WO9500651-A1.  
XX PD 05-JAN-1995.  
XX PF 20-JUN-1994; 94WO-IB00166.  
XX PR 18-JUN-1993; 93DK-0000725.  
XX PA (SYMB-) SYMBICOM AB.  
XX PI Eglerud T, Hansson L;  
XX DR WPI; 1995-052088/07.  
XX DR N-PSDB; AAO81203.  
XX PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme  
XX PT - and related vectors, transformed cells and polypeptides,  
XX PT useful for treating skin disorders, e.g. acne or psoriasis, and  
XX PT for identification of specific inhibitors.  
XX PS Disclosure; Page 97; 137pp; English.  
XX CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic  
XX CC and skin care products, especially to treat and prevent acne,  
XX CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or  
XX CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is  
XX CC produced recombinantly following mammal, insect, plant, or  
XX CC microorganism transformation with Plasmid pS507.  
XX CC (Updated on 25-MAR-2003 to correct PN field.)  
XX SQ Sequence 253 AA;  
XX  
Query Match 36.7%; Score 549; DB 16; Length 253;  
Best Local Similarity 40.9%; Pred. No. 1e-43;  
Matches 108; Conservative 45; Mismatches 97; Indels 14; Gaps 4;  
QY 14 ARAALKLPLMLMAQMAEAALLPONDTRLDPEAYGAPCARGSQPMQVSLFNGLSFHCAG 73  
DB 2 ARSLILPLQILLLSLALETRGEBAGQDKTID---GAPCARSHPMQVALLSNGQLHCAG 57  
QY 74 VLVDQSWVLTAAHCGNKPMLARVGDHLLLLQGEQLRTTRSVVHKYHQSGSPILPRRT 133  
DB 58 VLVNERVWLTAAHCGNMEYTVHLGSDTLGDRRAQRI-KASKSFRHFGY-----STQT 108  
QY 134 DEHDLMLKLARPVVPGPRVRLALQPYRCAGQGDQCVAGMGTTAARRYKNGKGLTCSGI 193  
DB 109 HVDLMLVLTNSQARLSMWKRVLPSCRPEPGITCTVSGMGTTSPDVTFFSDLMCVDV 168  
QY 194 TLSPKECEVFYFGVYVNNMICAGL-DRGQDPQSDSGSPVLVDETLGILSGVYPCGS 252  
DB 169 KLSFDCTKVTYVDLLENSMLCAGIDPSKKNACNBSGGSLVCRGTLGSLVSGTFFCGQ 228  
QY 253 AQHPAVYTIQCKTMSINKVIRSN 276  
DB 229 PNDPGVYTVQCKFTKWINDTMKKH 252

RESULT 13  
AAW05383  
ID AAW05383 standard; Protein; 253 AA.  
XX  
XX AAW05383;  
XX  
XX 31-DEC-1996 (first entry)  
XX  
XX

DE Human amyloid precursor protein protease.  
XX KW Amyloid precursor protein protease; Alzheimer's disease; diagnosis;  
XX KM therapy.  
XX OS Homo sapiens.  
XX PN WO9631122-A1.  
XX PD 10-OCT-1996.  
XX PF 02-APR-1996; 96WO-US04294.  
XX PR 04-APR-1995; 95US-0416257.  
XX PA (ELIL ) LILLY & CO ELI.  
XX PI Dixon EP, Johnstone EM, Little SP;  
XX DR WPI; 1996-464694/46.  
XX DR N-PSDB; AAT39783.  
XX PT New isolated human amyloid precursor protein protease - used to  
XX PT develop prods. for the treatment or diagnosis of associated  
XX PT conditions, esp. Alzheimer's disease  
XX PS Claim 1, Page 44-45; 55pp; English.  
XX CC Human amyloid precursor protein protease (AAW05383) is involved in  
XX CC the processing or clearance of amyloid precursor protein to form  
XX CC beta-amyloid peptide. Its amino acid sequence was deduced from  
XX CC a cDNA clone (AAT39783) obtd. from a human lung library. Recombinant  
XX CC protease can be produced in transformed or transfected prokaryotic  
XX CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is  
XX CC used to develop products for the design and testing of cpds. useful  
XX CC for treating or preventing conditions associated with beta-amyloid  
XX CC peptide, esp. Alzheimer's disease.  
XX SQ Sequence 253 AA;  
XX  
Query Match 36.7%; Score 549; DB 17; Length 253;  
Best Local Similarity 40.9%; Pred. No. 1e-43;  
Matches 108; Conservative 45; Mismatches 97; Indels 14; Gaps 4;  
QY 14 ARAALKLPLMLMAQMAEAALLPONDTRLDPEAYGAPCARGSQPMQVSLFNGLSFHCAG 73  
DB 2 ARSLILPLQILLLSLALETRGEBAGQDKTID---GAPCARSHPMQVALLSNGQLHCAG 57  
QY 74 VLVDQSWVLTAAHCGNKPMLARVGDHLLLLQGEQLRTTRSVVHKYHQSGSPILPRRT 133  
DB 58 VLVNERVWLTAAHCGNMEYTVHLGSDTLGDRRAQRI-KASKSFRHFGY-----STQT 108  
QY 134 DEHDLMLKLARPVVPGPRVRLALQPYRCAGQGDQCVAGMGTTAARRYKNGKGLTCSGI 193  
DB 109 HVDLMLVLTNSQARLSMWKRVLPSCRPEPGITCTVSGMGTTSPDVTFFSDLMCVDV 168  
QY 194 TLSPKECEVFYFGVYVNNMICAGL-DRGQDPQSDSGSPVLVDETLGILSGVYPCGS 252  
DB 169 KLSFDCTKVTYVDLLENSMLCAGIDPSKKNACNBSGGSLVCRGTLGSLVSGTFFCGQ 228  
QY 253 AQHPAVYTIQCKTMSINKVIRSN 276  
DB 229 PNDPGVYTVQCKFTKWINDTMKKH 252

RESULT 14  
ABB84406  
ID ABB84406 standard; Protein; 253 AA.  
XX  
XX ABB84406;  
XX  
XX 08-NOV-2002 (first entry)  
XX  
XX

DE Human SCCE protein.  
 XX  
 XX SCCE, human; stratum corneum chymotryptic enzyme; kallikrein 7;  
 KW serine protease; transgenic mammal; skin; skin disease; skin cancer;  
 KM hyperkeratosis; acanthosis; epidermal inflammation; dermal inflammation;  
 XX pruritus; atopic dermatitis; eczema; acne; itch; KLK7; enzyme.  
 OS  
 XX Homo sapiens.  
 PN WO200262135-A2.  
 PD 15-AUG-2002.  
 XX  
 XX 08-FEB-2002; 2002MO-IB01300.  
 XX  
 XX 09-FEB-2001; 2001CA-2332655.  
 PR 09-FEB-2001; 2001DK-0000218.  
 XX  
 XX (EGEL/) EGELRUD T.  
 PA (HANS/) HANSSON L.  
 XX  
 XX Egelrud T, Hansson L;  
 DR WPI; 2002-643380/69.  
 DR N-PSDB; ABQ76226.  
 XX  
 PT Transgenic mammal or its embryo useful as model for human disease, has  
 PT heterologous nucleotide sequence coding for stratum corneum  
 PT chymotryptic enzyme operably linked to promoter that drives its  
 PT expression in skin -  
 XX  
 PS Claim 10; Page 58-59; 74pp; English.  
 XX  
 CC This invention describes a novel non-human transgenic mammal or mammalian  
 CC embryo having integrated within its genome, a heterologous nucleotide  
 CC sequence comprising at least a significant part of a nucleotide sequence  
 CC coding for a stratum corneum chymotryptic enzyme (SCCE) or its variant,  
 CC operably linked to a promoter that drives expression of heterologous scce  
 CC or its variant in skin. The product of the invention is useful as a model  
 CC for the study of disease with the aim of improving treatment, to relieve  
 CC or ameliorate a pathogenic condition, for development or testing of a  
 CC cosmetic or a pharmaceutical formulation, and for the development of a  
 CC diagnostic method. It can also be used as a model for a skin disease or  
 CC skin cancer. The invention is also useful for screening or identifying a  
 CC compound or composition effective for the prevention or treatment of an  
 CC abnormal or unwanted phenotype, and for screening or identifying a  
 CC compound or composition effective for the prevention or treatment of  
 CC inflammatory skin diseases selected from diseases consisting of epidermal  
 CC hyperkeratosis, acanthosis, eczema, acne and inherited skin diseases  
 CC pruritus, atopic dermatitis, eczema, acne and inherited skin diseases  
 CC as a model for further studies of itch mechanisms and the testing of  
 CC potential compounds and compositions for relieve of various skin diseases  
 CC where itch is a component. This sequence represents the human stratum  
 CC corneum chymotryptic enzyme, SCCE which is a serine protease synonymous  
 CC with human Kallikrein 7 (KLK7) and is used in the development of the  
 CC transgenic mammals described in the invention.  
 XX  
 SQ Sequence 253 AA;  
 Query Match 36.7%; Score 549; DB 23; Length 253;  
 Best Local Similarity 40.9%; Pred. No. 1e-43; Indels 14; Gaps 4;  
 Matches 108; Conservative 45; Mismatches 97;  
 QY 14 ARAIAKLPLIMQIMAAEALPQNDTRLDPEAYGAPCARSGOPWVSFNGLSFRCAG 73  
 DB 2 ARSLLPLQILLBLAETGEBAQGXKIID---GAPCARSGHPWQVALLSGWQLHCGG 57  
 QY 74 VLVDQSVTLTAAGCKNPEYARVDDHLLHQSQAKRTTSVTHPKYHSGSGTLPKRT 133  
 DB 58 VLVNERVWLTAAHCKNMEYTVLHGSDDLGRARRRI-KASGSFHPG-----STQT 108  
 QY 134 DEHLMILKLARPVVGPFRVALQIPYKCAQPGQCQVAGWGTAAARRVXNKGLTCSI 193

DB 109 HNDMLVKNLSQARLSMTVKVRLPESRCPEPGTTCVSGWGTTS PDVTFPSDLMCDV 168  
 QY 194 TILSKKEEVYVPGVVTNNMTCAGL-DRGDPQCSGSGPVCDETLOGIISWGYPCGS 252  
 DB 169 KLISPDCTKYKYLLESMILCAGIPDSKQACNGDSGLVCHGTGLVSGWTFPCGQ 228  
 QY 253 AQHPAVYTOICKYMSINKVIRSN 276  
 DB 229 PNDGVTYOVCFKTMINDYMKR 252  
 RESULT 15  
 ABB84421  
 ID ABB84421 standard; Peptide; 253 AA.  
 XX  
 XX ABB84421;  
 XX  
 XX 08-NOV-2002 (first entry)  
 XX  
 XX Human SCCE protein N-terminal fragment SEQ ID 48.  
 XX  
 XX SCCE, human; stratum corneum chymotryptic enzyme; kallikrein 7;  
 KW serine protease; transgenic mammal; skin; skin disease; skin cancer;  
 KM hyperkeratosis; acanthosis; epidermal inflammation; dermal inflammation;  
 KW pruritus; atopic dermatitis; eczema; acne; itch; KLK7; enzyme.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200262135-A2.  
 PD 15-AUG-2002.  
 XX  
 XX 08-FEB-2002; 2002MO-IB01300.  
 XX  
 XX 09-FEB-2001; 2001CA-2332655.  
 PR 09-FEB-2001; 2001DK-0000218.  
 XX  
 XX (EGEL/) EGELRUD T.  
 PA (HANS/) HANSSON L.  
 XX  
 XX Egelrud T, Hansson L;  
 DR WPI; 2002-643380/69.  
 DR  
 XX  
 PT Transgenic mammal or its embryo useful as model for human disease, has  
 PT heterologous nucleotide sequence coding for stratum corneum  
 PT chymotryptic enzyme operably linked to promoter that drives its  
 PT expression in skin -  
 XX  
 PS Example 6; Page 37; 74pp; English.  
 XX  
 CC This invention describes a novel non-human transgenic mammal or mammalian  
 CC embryo having integrated within its genome, a heterologous nucleotide  
 CC sequence comprising at least a significant part of a nucleotide sequence  
 CC coding for a stratum corneum chymotryptic enzyme (SCCE) or its variant,  
 CC operably linked to a promoter that drives expression of heterologous scce  
 CC or its variant in skin. The product of the invention is useful as a model  
 CC for the study of disease with the aim of improving treatment, to relieve  
 CC or ameliorate a pathogenic condition, for development or testing of a  
 CC cosmetic or a pharmaceutical formulation, and for the development of a  
 CC diagnostic method. It can also be used as a model for a skin disease or  
 CC skin cancer. The invention is also useful for screening or identifying a  
 CC compound or composition effective for the prevention or treatment of an  
 CC abnormal or unwanted phenotype, and for screening or identifying a  
 CC compound or composition effective for the prevention or treatment of  
 CC inflammatory skin diseases selected from diseases consisting of epidermal  
 CC hyperkeratosis, acanthosis, eczema, acne and inherited skin diseases  
 CC pruritus, atopic dermatitis, eczema, acne and inherited skin diseases  
 CC with epidermal hyperkeratosis. The mammal of the invention is also useful  
 CC as a model for further studies of itch mechanisms and the testing of  
 CC potential compounds and compositions for relieve of various skin diseases  
 CC where itch is a component. This sequence represents the N-terminal

CC fragment of the human stratum corneum chymotryptic enzyme, SCCE  
CC synonymous with human kallikrein 7 (KLK7), used in the development of the  
CC transgenic mammals described in the invention.  
xx